UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Roth, L., Cipora, K., Overlander, A. T., Nuerk, H.-C., & Reips, U.-D. (2026). Shape of SNARC: How Task-Dependent Are Spatial-Numerical Associations? A Highly Powered Online Experiment. *Collabra: Psychology*, *12*(1). https://doi.org/10.1525/collabra.170661

Cognitive Psychology

Shape of SNARC: How Task-Dependent Are Spatial-Numerical Associations? A Highly Powered Online Experiment

Lilly Roth[1], Krzysztof Cipora[2,3], Annika Tave Overlander[4], Hans-Christoph Nuerk[1,5,6], Ulf-Dietrich Reips[4][a]

[1] Department of Psychology, University of Tuebingen, Tuebingen, Germany, [2] Centre for Mathematical Cognition, Loughborough University, Loughborough, UK, [3] Copernicus Center for Interdisciplinary Studies, Jagiellonian University, Kraków, Poland, [4] Department of Psychology, University of Konstanz, Konstanz, Germany, [5] LEAD Graduate School & Research Network, University of Tuebingen, Tuebingen, Germany, [6] German Center for Mental Health (DZPG)

Keywords: spatial-numerical associations, SNARC effect, magnitude classification, parity judgment, task dependency

https://doi.org/10.1525/collabra.170661

Collabra: Psychology

Vol. 12, Issue 1, 2026

Spatial-Numerical Associations are fundamental to numerical cognition. The Spatial-Numerical Association of Response Codes (SNARC) effect denotes faster responses to small/large numbers with the left/right, respectively (Dehaene et al., 1993). It is typically measured in *magnitude classification* (MC: smaller vs. larger than 5) or *parity judgment* (PJ: odd vs. even) with numbers 1–9 (excluding 5). These tasks differ in magnitude and parity processing, compatibility effects, and other aspects. In this highly powered online experiment, data from 611 participants in a within-subjects design demonstrated a rather categorical MC-SNARC (i.e., similar left-hand advantage for 1–4, and similar right-hand advantage for 6–9) and a continuous PJ-SNARC (i.e., right-hand advantage increasing with magnitude). Accordingly, the PJ-SNARC can be modeled using linear regression, while the MC-SNARC requires a categorical or mixed analysis. Explained variance and split-half reliability were higher for the MC-SNARC than for the PJ-SNARC. We replicated the Numerical Size Effect (slower responses for larger numbers; Brysbaert, 1995) in MC and PJ, the Numerical Distance Effect (faster responses for larger numerical distance between stimulus and reference; Moyer & Landauer, 1967) in MC, and the Odd effect in PJ (faster responses to even numbers; Hines, 1990; potentially masked by the Numerical Distance Effect in MC), but not the MARC effect (faster responses to odd/even with the left/right, respectively; Nuerk et al., 2004). The MC-SNARC and PJ-SNARC correlated slightly positively (r = 0.17), suggesting partly common underlying magnitude representations. Altogether, the SNARC effect appears highly task-dependent and should be modeled differently depending on the task.

Spatial-numerical associations (SNAs) belong to the fundaments of numerical cognition (Fischer & Shaki, 2014; Toomarian & Hubbard, 2018). They have been implicated as an important representation (e.g., Dehaene et al., 2003) and as means to foster numerical and arithmetic learning (Booth et al., 2008; Dackermann et al., 2017, for an overview of embodied spatial-numerical learning). SNAs can be divided into spatial-extensional SNAs, where a particular number or magnitude is related to a physical extension (i.e., larger number to larger extensions), and directional SNAs, where a particular number is associated with a particular location in space (Patro et al., 2014). Both SNAs are important and seem to be highly dependent on the experimental situation or task (Cipora, Patro, et al., 2018). Understanding such situational dependencies is key to understanding SNAs and their relation to mathematics as such (Cipora, He, et al., 2020).

SNAs can refer to explicit or implicit associations of different characteristics of numbers (e.g., cardinality, ordinality, parity) with different aspects of space, namely directions or extensions (Cipora, Schroeder, et al., 2018; Cipora, Haman, et al., 2020; Patro et al., 2014). For instance, the MARC effect (Linguistic Markedness of Response Codes; Nuerk et al., 2004) reflects the association between parity (odd/even numbers) and direction (the left/right side), respectively. The hallmark directional SNA, however, is the Spatial-Numerical Association of Response Codes (SNARC) effect, which denotes that – at least in left-to-right reading cultures – participants respond faster to small/large magnitude numbers on the left/right side, respectively (Dehaene et al., 1993). SNAs are claimed to reflect implicit and explicit mental representations of numbers and processes operating on them (Cipora, Haman, et al., 2020). The tendency to map characteristics of numbers onto space is

a Corresponding author: reips@uni-konstanz.de

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

considered one of the basic traits of human cognition (Cipora, Patro, et al., 2018). The SNARC effect has been replicated with stimuli in different modalities and notations (e.g., visual Arabic numerals, visual number words, auditory number words, visual dice patterns; Nuerk et al., 2004; Nuerk, Wood, et al., 2005) and in different response setups (e.g., manual responses, pedal responses, saccadic eye movements; Schwarz & Keus, 2004; Schwarz & Müller, 2006), offline as well as online (Cipora et al., 2019; Roth et al., 2024). The association of number magnitude and space therefore seems to be highly robust and generalizable across many settings, even though many situational modulations have been described (Cipora et al., 2018).

Importantly, the SNARC effect arises in several tasks, which – as we outline in detail below – have major conceptual differences in the underlying semantic features of the numbers being processed. Two tasks that inquire about semantic numerical attributes of the digits/numbers themselves are by far the most frequently used to investigate the SNARC effect: the *magnitude classification* (MC) task and the *parity judgment* (PJ) task (see Table 2 in Wood et al., 2008). In MC, participants judge whether numbers are smaller or larger than a reference number[1]. In PJ, participants judge whether numbers are odd or even. Although some studies have used other kinds of stimuli (e.g., dice patterns or number words in Nuerk, Wood, et al., 2005, and multi-digit numbers in Tlauka, 2002; Weis et al., 2018), single-digit (i.e., Arabic or Hindu-Arabic) numbers are most used in this task. In both tasks, the instruction is to respond as quickly and as accurately as possible with a left- or right-hand key to numbers presented centrally on a computer screen. Typically, symbolic[2] numbers from 1 to 9 (excluding 5) are used as the stimulus set, with number 5 serving as the reference number in MC (Wood et al., 2008). In both tasks, the response-to-key assignment is flipped in the middle of the experiment, so that both left- and right-hand responses are given for each number.

In the following, we first outline conceptual differences between MC and PJ (e.g., the relevance of number magnitude and number parity, SNARC and MARC compatibility, and the strength of the SNARC effect in MC and PJ) and a potential correlation of the SNARC effect between tasks. Most crucially, we then discuss previously observed shape differences, and reasons for a categorical SNARC effect in MC. Additionally, we introduce the potential influence of task order on the SNARC effect. At the end of this introduction, we give an overview of all replication checks (see Table 1), hypotheses (see Table 2), and exploratory tests (see Table 3) we aimed at testing in the current study.

Conceptual Differences Between MC and PJ

At first sight, apart from the instructions, the MC and PJ tasks seem to be similar: A semantic feature of numbers (i.e., magnitude or parity) is to be categorized. In the current study, we used the most common bimanual computerized setup with symbolic numbers from 1 to 9 (excluding 5) described above. One might assume that the required cognitive processes, the responses given by participants, and the arising spatial mapping of number magnitude are similar in both tasks. However, as we shall see, the SNARC effects in MC and PJ (called *MC-SNARC* and *PJ-SNARC* in the remainder of this article) differ. In the following, we describe conceptual differences between MC and PJ to shed light on reasons for the different SNARC effects.

Relevance of Number Magnitude and Number Parity

Most obviously, number magnitude is directly task-relevant to MC but not to PJ. This leads to an important difference: Tzelgov et al. (2015) distinguish between intentional and automatic processing, where the latter means processing without conscious monitoring. Automatic processing can be measured in tasks where the process in question is not part of the task requirements (Tzelgov, 1997). In contrast, intentional processing is supposed to reflect the task requirements.

For participants to show the SNARC effect, two representations must be activated, namely the *magnitude* (e.g., two) or *ordinality* (e.g., the second)[3] of a number and its directional association with space (Cipora, He, et al., 2020). Importantly, when judging parity, neither the processing of magnitude nor of its directional association with space is task-relevant and intentional. As a result, the PJ-SNARC effect is commonly regarded as an indicator of automatic number magnitude processing in humans. It also supports the notion that single-digit numbers function as primitives in Western cultures, such that their meanings can be effortlessly and holistically retrieved from memory without additional processing (Tzelgov et al., 2015). This is fundamentally different for MC, where the processing of magnitude is intentional, as it is task-relevant. Therefore, the MC-SNARC only shows that the directional association with space is automatic (as it is not needed for the response), but not that number-magnitude processing is automatic.

Importantly, the two tasks differ not only regarding task-relevance of number magnitude but also in the task-relevance of number parity. Parity is task-relevant in PJ while being task-irrelevant in MC. Apart from the SNARC effect,

1 In the current article, as in Wood et al. (2008), tasks in which presented numbers are to be compared with a fixed reference number (e.g., comparing whether a presented number between 1 and 9 excluding 5 is smaller or larger than 5) are referred to as magnitude *classification*. In contrast, tasks in which the reference number is not fixed but varies between trials are referred to as magnitude *comparison*.

2 Symbolic numbers consist of symbols, such as digits, whereas non-symbolic numbers consist of quantities, such as dot arrays.

3 A vivid discussion exists in the field on whether it is magnitude or ordinality that drives the SNARC effect (Casasanto & Pitt, 2019; Prpic et al., 2016; Van Dijck et al., 2014; and see Introduction section "Further differences between MC and PJ"). The design of the present study does not permit to disentangle their roles, and we refer to magnitude in the manuscript for simplicity reasons.

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

a phenomenon to be considered in the current study is the MARC effect (Linguistic Markedness of Response Codes; Nuerk et al., 2004), which is typically observed in PJ but not in MC (Replication Check 2; see ordered list of replication checks in the Introduction section "The current study"). The MARC effect reflects faster responses to odd/even numbers on the left/right side, respectively. In a similar vein as for the SNARC effect, there are two prerequisites for the MARC effect: the processing of parity and its directional association with space. Crucially, the processing of number parity is less automatic than the processing of number magnitude (Roth, Caffier, Reips, Cipora, et al., 2025) and is more consistently found when using number words rather than Arabic digits (Nuerk et al., 2004; Roettger & Domahs, 2015). The processing of number parity seems not to be automatic, as typically no MC-MARC is found (Cipora, 2014; Deng et al., 2018). In contrast, the directional association of parity with space can be considered automatic, because it is not required in PJ, yet a PJ-MARC can typically be observed.

Further, response times (RT) increase when numerical magnitude increases, which is referred to as the Numerical Size Effect (Brysbaert, 1995). We expect the effect to arise in both tasks (Hypothesis 3a; see ordered list of hypotheses in the Introduction section "The current study"), although it has mainly been demonstrated for two-digit numbers (Brysbaert, 1995) and single-digit numbers were used in the present study (as in Hohol et al., 2020). Moreover, we expect it to be stronger in MC than in PJ (Hypothesis 3b) because numerical size (i.e., number magnitude) is only task-relevant in MC. Moreover, RT decrease with increasing numerical distance between the stimulus and the reference number in MC, which is referred to as the Numerical Distance Effect (Moyer & Landauer, 1967). We expect the effect to arise in MC (Replication Check 4), but it cannot arise in PJ because there is no criterion that numbers are compared to.

SNARC and MARC Compatibility

Importantly, the SNARC and MARC effects can also be considered *compatibility effects* (e.g., see cognitive-control account for the SNARC effect by Zhang et al., 2022). Crucially, MC and PJ differ with respect to compatibility. In a typical PJ task with two blocks, one block is MARC-compatible (i.e., when the instruction is to respond to odd/even numbers with the left-/right-hand key, respectively) and the other block is MARC-incompatible. At the same time, within each block, half of the trials are SNARC-compatible (i.e., when the response to the parity of small/large numbers is assigned to the left/right, respectively) and half of the trials are SNARC-incompatible. On the contrary, in a typical MC task with two blocks, one block is SNARC-compatible (i.e., when participants are asked to respond to numbers that are smaller/larger than a reference number with the left-/right-hand key, respectively) and the other is SNARC-incompatible. At the same time, within each block, half of the trials are MARC-compatible (i.e., when the response to the magnitude of odd/even numbers is assigned to the left/right, respectively) and half of the trials are MARC-incompatible. SNARC-compatible and -incompatible trials alternating within blocks, as is the case in PJ, can elicit Gratton effects (Gratton et al., 1992). In line with this, Pfister et al. (2013) reported reduced SNARC effects after SNARC-incompatible compared to after SNARC-compatible trials in PJ. In a typical MC task with only two blocks, where SNARC-compatibility is grouped by block, such trial-to-trial effects cannot occur.

Apart from trial-to-trial compatibility effects, there can be compatibility effects depending on block order. Van Galen and Reitsma (2008) observed a stronger MC-SNARC in children who completed the SNARC-compatible followed by a SNARC-incompatible block than in children who were administered the reverse order, whereas Bulut et al. (2025) observed the opposite effect in one of three tested adult samples, but no effect in the two remaining adult samples. No effect of block order on the PJ-SNARC has been found (Bulut et al., 2025; Cipora, van Dijck, et al., 2019; Moorkens et al., 2025; Roth et al., 2024), where each block is either MARC-compatible or MARC-incompatible, while it consists of half SNARC-compatible and half SNARC-incompatible trials. For the MARC effect, influences of block order have been found in both directions. In a previous study, we found a stronger MARC effect in PJ with the MARC-incompatible-compatible order compared to the MARC-compatible-incompatible order (block-wise compatibility order differing between-subjects, Cipora, van Dijck, et al., 2020). In another previous study, we found the reversed pattern (block-wise compatibility order differing within-subjects; Roth et al., 2024). The difference might be attributable to the design, and as block-wise compatibility order differed between-subjects in the current study like in the study by Cipora, van Dijck, et al. (2020), we expect the same pattern here. A stronger PJ-MARC in the MARC-incompatible-compatible order than in the MARC-compatible-incompatible order seems plausible: Participants need to familiarize themselves with PJ and overcome their natural odd-left and even-right association in the first block, while they are already familiar with PJ and can respond in line with their natural odd-left and even-right association in the second block. Participants therefore have two reasons to be slower in the first block, and the difference (i.e., the MARC effect) between blocks is therefore especially strong in this block order. We exploratorily investigated compatibility-order effects in both tasks in the current study (Exploratory 2; see ordered list of exploratory tests in the Introduction section "The current study").

Strength of the SNARC Effect in MC and PJ

As outlined above, the processing of number magnitude is highly automatized and single-digit numbers can therefore be considered as primitives in Western cultures' numerical cognition (Tzelgov et al., 2015). In contrast, the processing of number parity might be less automatized when only a semantic number feature other than parity is task-relevant (no MC-MARC found by Cipora, 2014, and Deng et al., 2018). Note that evidence has been found for both the MARC effect and the Odd effect (i.e., slower responses to odd than to even numbers, Hines, 1990) when

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

only non-semantic features of numbers were judged (e.g., font color; Roth, Caffier, Reips, Cipora, et al., 2025), reflecting automatic parity processing to some extent, but the evidence was only weak. Also, number magnitude is more often relevant in daily life than number parity. Hence, the processing of magnitude seems to be more straightforward than the processing of parity. In line with this, average responses are typically faster in MC than in PJ (Kiesel et al., 2007; Saeki & Saito, 2009; descriptively also observed by Fattorini et al., 2015; Fitousi et al., 2009; Gevers, Verguts, et al., 2006; Ito & Hatta, 2004, see also Wood et al., 2008, for a meta-analysis), which we expected to find in the current study as well (Replication Check 3).

The processing of magnitude being explicitly required in MC, but not in PJ, might elicit a stronger spatial mapping in MC than in PJ. In line with this assumption, the MC-SNARC has been found to be stronger than the PJ-SNARC (Bae et al., 2009; Cheung et al., 2015; Fitousi et al., 2009; Van Dijck et al., 2009). On the other hand, judgments of number magnitude are automatic (Tzelgov et al., 2015) and therefore naturally faster than judgments of number parity. At the same time, the SNARC effect is typically stronger in slower responses in both MC and PJ, both within participants and on the sample level (see Table ST4 in Supplementary Materials in Cipora, Soltanlou, et al., 2019; see Table 3 in Didino et al., 2019; see Figure 6 in Gevers, Verguts, et al., 2006; see Figure 4 in Kopiske et al., 2016). In contrast to the reasoning above, this would lead to the opposite prediction of the PJ-SNARC being stronger than the MC-SNARC, which has been observed by Georges et al. (2017), Gevers, Verguts, et al. (2006), and Ito and Hatta (2004). No difference between the MC-SNARC and the PJ-SNARC was found by Didino et al. (2019) in an independent-samples *t*-test. Taken together, (i) easier and possibly deeper processing of magnitude in MC than in PJ should lead to a greater SNA in MC, and (ii) longer response times in PJ than in MC should lead to a greater SNA in PJ. Both opposing processes seem to be valid and there is no clear picture in the literature. It remains unclear whether the SNARC effect differs in size between tasks, and we therefore look at this in Exploratory 6.

Further Differences Between MC and PJ

Several more differences may exist between MC and PJ. First, the MC-SNARC seems to more strongly involve visuospatial working memory, the PJ-SNARC seems to rely more on verbal working memory (Deng et al., 2017; Herrera et al., 2008; Van Dijck et al., 2009). Second, the MC-SNARC and the PJ-SNARC might arise at different processing stages (Basso Moro et al., 2018; Xiang et al., 2022). Third, cognitive mechanisms underlying the MC-SNARC and the PJ-SNARC might differ. Namely, Prpic et al. (2016) claim that ordinality drives the SNARC effect in *direct* tasks (e.g., MC, where magnitude or ordinality is response-relevant), whereas magnitude underlies in *indirect* tasks (e.g., PJ, where magnitude or ordinality is response-irrelevant). Note that Casasanto and Pitt (2019) claim that only ordinality is crucial for both direct and indirect tasks, and that Koch et al. (2023) show that order- and magnitude-related mechanisms are not mutually exclusive. Looking into these differences between the MC-SNARC and the PJ-SNARC is beyond the scope of the current study; however, the current study provides an improved understanding of the two tasks and thereby lays the groundwork for further investigation.

In summary, several conceptual differences exist between MC and PJ, concerning the task-relevance of number magnitude and number parity, the compatibility of the response-to-key assignment with the SNARC and MARC effects, arising numerical-cognition effects, and underlying cognitive mechanisms. However, both tasks elicit a SNARC effect, and the presence and strength of the MC- and PJ-SNARC in the same participants might be related to one another, as discussed in the next section.

Correlation Between the MC- and PJ-SNARC

After having described the similarities and differences of MC and PJ, the question arises whether the MC-SNARC and the PJ-SNARC are correlated. However, several factors might lead to a null correlation. First, there seem to be high fluctuations in the SNARC effect over time (Roth et al., 2024) that limit the maximum correlation that can be detected. Second, the test-retest reliability of the SNARC effect has been found to be poor for both MC and PJ (correlations $0.22 < r < 0.41$; Cipora & Göbel, 2013; Georges et al., 2013; Hedge et al., 2018; Viarouge et al., 2014). The lower the test-retest reliabilities of the MC-SNARC and PJ-SNARC, the lower is also the maximally observable correlation between the two effects. Third, the split-half reliability of the SNARC effect has been found to be poor for PJ at least in some studies (correlations $0.43 < r < 0.96$; for an overview, see Table 1 in Cipora, van Dijck, et al., 2019), although it seemed higher for MC in one study ($r = .89$; Hohol et al., 2020, see Supplementary Material 3). To conclude, both properties of the SNA construct and its operationalization in experimental tasks influence whether a correlation between the MC-SNARC and the PJ-SNARC is found. Possible reasons for a null finding could be low intraindividual stability, low reliability, or low internal consistency, whereas we would interpret a high correlation between the MC-SNARC and the PJ-SNARC as both MC and PJ reliably measuring the same underlying theoretical construct.

Several previous studies did not find any correlation (correlations with 95% confidence intervals and *p*-values: $r = -.02$ [-0.19, 0.15] and $p = .822$ for Germans; $r = -.08$ [-0.26, 0.10] and $p = .386$ for Turks; $r = .10$ [-0.13, 0.32] and $p = .402$ for Iranians, in Bulut et al., 2025; $r = .09$ [-0.18, 0.35] and $p = .513$ in Cipora, 2014; $r = .06$ [-0.30, 0.40] and $p = .744$ in Didino et al., 2019; $r = .18$ [-0.07, 0.42] and $p = .18$ in Fattorini et al., 2015; $r = .20$ [-0.01, 0.39] and $p = .07$ in Georges et al., 2017). To our knowledge, a statistically significant correlation has only been reported by Cheung et al. (2015; $r = .25$, $p < .05$) and by Cipora (2014; $r = .50$, $p < .01$, but only in a unimanual setup). Note that an existing weak correlation between the MC-SNARC and the PJ-SNARC despite limiting factors such as low intraindividual stability, low reliability, or low internal consistency would only be detectable in large samples. To be able to detect a potential correlation between the MC-SNARC and the PJ-SNARC,

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

we administered both tasks to a large sample in a within-subjects design. This enabled us to test the correlation between the MC-SNARC and the PJ-SNARC in Exploratory 7 with high statistical power.

Different Shapes of the SNARC Effect

After having outlined the differences between MC and PJ and after having discussed the potential correlation between the SNARC effect in these two tasks, it is important to note that the shape of the MC-SNARC seems to be categorical, while the shape of the PJ-SNARC seems to be continuous (Wood et al., 2008). However, the MC-SNARC is often modelled as a continuous phenomenon, just as the PJ-SNARC, as described in the following. The main aim of the current study is to thoroughly investigate the SNARC effect in these two most widely used tasks and to find out how to best statistically model the SNARC effect.

The SNARC effect is usually calculated by subtracting the mean RT with the left hand from those with the right hand for each number and regressing these differences (dRT) on magnitude as a continuous predictor in both MC and PJ. A negative regression slope reflects the increasing right-hand advantage for larger numbers and therefore the SNARC effect. To investigate whether the effect is present on group level, regression slopes (one per participant) are then tested against zero in a one-sample *t*-test (repeated-measures regression, adapted by Fias et al., 1996, based on Lorch & Myers, 1990). Importantly, this analysis method is only suitable for a continuous SNARC effect, reflecting a constant increase in right-hand advantage (reflected by a constant decrease in dRT) per increase of magnitude. Hence, when participants judge whether numbers in the typically used stimulus set from 1 to 9 (excluding 5) are odd or even, the spatial mapping of extreme magnitudes such as 1 and 9 is stronger than for magnitudes closer to the mid of the stimulus set such as 4 and 6. In other words, the association with the left side is stronger for the very small number 1 than for the slightly small number 4. While the PJ-SNARC is linear, the MC-SNARC is typically categorical (e.g., Gevers, Verguts, et al., 2006), especially in adults (Van Galen & Reitsma, 2008): In a typical MC task, when averaging across all participants, responses are *equally* faster with the left hand to numbers from 1 to 4 and *equally* faster with the right hand to numbers from 6 to 9. Therefore, a stepwise model reflects the MC-SNARC better than a continuous model (as reflected by a better model fit in terms of a higher proportion of explained variance). The use of a categorical instead of a linear function for quantifying the MC-SNARC would increase the model fit at the participant level and thereby likely also the precision of the effect size estimate at the sample level.

Nevertheless, a linear predictor in the regression of dRT on number magnitude remained a frequently used analysis of the MC-SNARC (Bachot et al., 2005; Bae et al., 2009; Bull et al., 2005; Cheung et al., 2015; Deng et al., 2017; Han et al., 2017; Herrera et al., 2008; Hoffmann et al., 2013; Hubbard et al., 2009; Ito & Hatta, 2004; Lohmann et al., 2018; Mourad & Leth-Steensen, 2017; Nathan et al., 2009; Pinto et al., 2021; Schiller et al., 2016; Shaki & Gevers, 2011; Van Dijck et al., 2009, 2012; Van Dijck & Doricchi, 2019; Van Galen & Reitsma, 2008; Weis et al., 2018). As the correlation between the linear predictor (i.e., 1, 2, 3, 4, 6, 7, 8, 9) and a categorical predictor (especially the most commonly assumed one, i.e., -0.5, -0.5, -0.5, -0.5, 0.5, 0.5, 0.5, 0.5) is extremely high (r = .913), the model with a linear magnitude predictor fits relatively well both to the continuously and categorically distributed dRT in MC and PJ (see top panel of Figure 1 in Bae et al., 2009, or Figure 1 in Nathan et al., 2009). In some studies, a two-way ANOVA including magnitude (small vs. large) and response side (left vs. right) as within-subjects factors has been used to quantify the MC-SNARC (Fattorini et al., 2015; Gevers, Verguts, et al., 2006; Herrera et al., 2008; Hoffmann et al., 2013; Nathan et al., 2009). However, compared to that approach, the repeated-measures regression approach has several advantages (Fias et al., 1996): First, the presence of the SNARC effect is judged by a main effect instead of an interaction effect, which allows a quantification of the size of the effect in milliseconds by the slope. Second, the presence or absence of a SNARC effect can be assessed for each participant individually. A repeated-measures regression with a categorical predictor for the MC-SNARC has only been used in few studies (Bulut et al., 2025; Cipora, 2014; Didino et al., 2019; Fitousi et al., 2009; Georges et al., 2017; Gevers, Verguts, et al., 2006; Hohol et al., 2020; Nathan et al., 2009; Nuerk, Wood, et al., 2005; Weis et al., 2018; Zorzi et al., 2012). For an overview of all mentioned studies including MC, see Table A1 in Appendix A. The categorical predictor was defined using 5 as a boundary (so that 1 to 4 are considered as "small" and 6 to 9 as "large") in all the named studies, and, if not specified otherwise, the same categorical predictor is meant in the current article (for alternative considerations, see below in Exploratory 4).

Unfortunately, the suitability of the linear and categorical predictors for dRT in MC with the stimulus set from 1 to 9 (excluding 5) was assessed by direct comparison in only a few studies. Fitousi et al. (2009) and Nathan et al. (2009) computed two separate regression models, one of which with a categorical and the other with a linear predictor, and in both studies the fit was higher with a categorical (R^2 = .904 in Fitousi et al., 2009; R^2 = .988 in Nathan et al., 2009) than with a linear predictor (R^2 = .775 in Fitousi et al., 2009; R^2 = .891 in Nathan et al., 2009). Similarly, Didino et al. (2019), Gevers, Verguts, et al. (2006) and Nuerk, Bauer, et al. (2005) ran regression analyses including both predictors, and only the categorical predictor turned out to be significant in all three studies. In a study with two-digit numbers where participants performed PJ and MC for the unit digit, Weis et al. (2018) also included both linear and categorical predictors for both unit and decade magnitude concurrently into one regression model. They found only a categorical predictor for units and decades to be significant in MC and only the linear predictors for units and decades to be significant in PJ, providing further evidence for a categorical MC-SNARC and a continuous PJ-SNARC. Importantly, because a linear model fits well even for the categorical MC-SNARC (e.g., see Fitousi et al., 2009; and Nathan et al., 2009), evidence for a better fit of a categorical model in MC can only

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

be achieved with sufficient power by using a large sample and a sufficient number of repetitions per experimental cell (resulting from the combination of each stimulus with each response hand per task). As outlined above, although a linear model fit to the MC-SNARC might be high in some studies, a categorical model seems to be more adequate. We expected to find a better fit of a categorical model in MC (Hypothesis 1) and of a linear model in PJ (Hypothesis 2).

Importantly, the boundary between "small" and "large" numbers is not necessarily 5 for everyone and might instead vary between individuals. A split of the full number interval into two halves (i.e., from 1 to 4 and from 6 to 9) seems plausible in MC, where the boundary of 5 is explicitly defined in the task instructions. However, especially in PJ, but potentially even in MC, some individuals might classify numbers into the categories "small" and "large" with a different boundary.[4] Crucially, an overall continuous SNARC effect could result from continuous patterns in most individuals, but also from averaging across categorical patterns differing between individuals. In this study, we therefore additionally determined the most likely categorical boundary for each participant separately. Subsequently, we investigated the shape of the SNARC effect (Hypotheses 1 and 2) once more, this time comparing the best fitting categorical model for each individual with the continuous model (Exploratory 4).

Explanations for the Categorical MC-SNARC Shape

The literature provides several explanations for the different shapes of the SNARC effect, depending on the task. First, numbers are typically roughly classified into small and large numbers (Banks et al., 1976; Tzelgov et al., 1992), which is sufficiently precise to perform MC and might lead to the categorical MC-SNARC (Fitousi et al., 2009; Gevers, Verguts, et al., 2006). In contrast, participants are not instructed to process number magnitude at all in PJ, and thus number magnitude processing is not intentional (i.e., slow conditional route according to the dual-route model by Gevers, Ratinckx, et al., 2006) but rather automatic (i.e., fast unconditional route)[5]. Automatic number magnitude processing seems to be more exact and is continuously mapped onto space in PJ. This explanation is in line with the polarity-correspondence account of the SNARC effect by Proctor and Cho (2006), as well as with the application of the markedness principle to number magnitude (Nuerk & Schroeder, 2024; Schroeder et al., 2017). According to these two theories, the SNARC effect arises because both *large* and *right* are associated with the positive or unmarked polarity and both *small* and *left* with the negative or marked polarity. Similarly, this explanation is compatible with the verbal-spatial account of the SNARC effect proposed by Gevers, Verguts, et al. (2006; see also Gevers et al., 2010), stating that verbal categories such as *small* vs. *large* and *left* vs. *right* are responsible for the SNARC effect. These accounts argue for an intermediate classification into small or large numbers (Santens & Gevers, 2008), and the polarities, markedness, or verbal labels are categorical rather than continuous (Bae et al., 2009), which explains the categorical MC-SNARC. Note that it is possible that the PJ-SNARC is linear in the beginning of the task and becomes categorical over the course of the task, so that the continuous shape shifts to a stepwise one. That is, participants might start classifying the stimuli into the two categories "small" and "large" in PJ as soon as they become familiar with the stimulus set because they might notice that the stimulus set consists of two single-digit number sequences (i.e., 1 to 4 and 6 to 9) separated by the missing number 5. We investigated this possibility in Exploratory 5.

Second, the Numerical Distance Effect might play a role for the MC-SNARC (Gevers, Verguts, et al., 2006). For instance, if the reference number is 5 in MC with the stimulus set from 1 to 9, numbers 4 and 6 need to be processed more intensely than numbers 1 and 9 to discriminate them from number 5 (Wood et al., 2008). Hence, responses are slowest for numbers 4 and 6 and fastest for numbers 1 and 9 with this stimulus set. In combination with the finding that the SNARC effect becomes stronger with increasing RT, the absolute values of dRT for number magnitudes that are close to the reference are larger than dRT predictions by the linear SNARC regression slope, resulting in a categorical shape (Didino et al., 2019; Georges et al., 2017; Gevers, Verguts, et al., 2006). Importantly, the Numerical Distance Effect demonstrates automatic processing of number magnitude (i.e., fast unconditional route), because it reflects performance differences in the discrimination between numbers and arises although the task instructions do not favor or disfavor the performance for specific stimuli. It thus does not build on the gross classification into smaller and larger numbers described in the previous paragraph (i.e., slow conditional route), but rather on the exact number magnitudes.

In summary, the rationale for a categorical instead of linear MC-SNARC is twofold: First, the intentional classification into small and large numbers is categorical in MC, and second, the interaction between the Numerical Distance Effect and the positive correlation between the SNARC effect and overall RT contributes to a step-wise shape. Because statistical models should correspond to scientific models as closely as possible (Westermann & Hager, 1983), the MC-SNARC should therefore be tested with a categorical predictor.

4 We wish to thank the stage-1 PCI-RR reviewer Peter Wühr for bringing this conceptual alternative to our awareness.

5 As outlined by the stage-1 PCI-RR reviewer Peter Wühr, predictions can be derived from the dual-route model. Importantly, both the automatic and the intentional route are activated in MC, whereas only the automatic route is activated in PJ. Thus, the SNARC effect should be stronger in MC than in PJ because it results from both routes instead of only one route. Moreover, a positive correlation of the MC- and PJ-SNARC can be assumed based on the dual-route model, since both effects are (at least partly) caused by the automatic route.

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Influence of Task Order on the SNARC Effect

The influence of task order (first MC and second PJ, or reversed) on the MC-SNARC and on the PJ-SNARC has been investigated in only a few studies. Didino et al. (2019) did not find any task-order effects on the SNARC effect. Fattorini et al. (2015) did not observe any task-order effect on the PJ-SNARC but found the MC-SNARC to be weaker after PJ than when MC was the first task. In most other studies that included the two tasks, effects of task order have either not been reported (Cheung et al., 2015; Gevers, Verguts, et al., 2006; Nuerk, Bauer, et al., 2005; Weis et al., 2018), or could not be calculated because task order was not counterbalanced (Bae et al., 2009; Cipora, 2014; Fitousi et al., 2009; Georges et al., 2017; Zorzi et al., 2012), or because different samples completed MC and PJ (Ito & Hatta, 2004; Van Dijck et al., 2009). To our knowledge, only Bulut et al. (2025; see Supplementary Materials) have tested the influence of task order, and they did not find an influence on the SNARC effect in any of the two tasks in any of three samples (130 German, 112 Turkish, and 75 Iranian participants). Two opposite theoretical predictions can be made. On the one hand, the SNARC effect might be stronger in each task if it is the second, because the processing of number magnitude and its spatial mapping should be stronger if they have already been activated in a previous task. On the other hand, the SNARC effect might be weaker in each task if it is the second, because RT typically decrease with practice and faster RT are typically associated with a weaker SNARC effect (note that both decreasing RT and a decreasing SNARC effect over time in PJ have been found by Roth et al., 2024). If both mechanisms were true, they might cancel out each other and make the influence of task order invisible. Hence, we could not make any directional prediction and investigated the potential influence of task order in Exploratory 1.

The Current Study

In this large-scale online study, we investigated whether the MC-SNARC is truly categorical (i.e., better described by a categorical number magnitude predictor) and the PJ-SNARC continuous (i.e., better described by a continuous number magnitude predictor). Evidence for this systematic difference would suggest that we should not call it *the* SNARC effect, but instead acknowledge that different SNARC effects exist, which are elicited depending on the task. It is crucial to shed light on this issue, because wrong measurements and interpretations of the SNARC effect can lead to misconceptions of SNAs. Another goal of the present study was to investigate the relationship between the SNARC effect(s) observed in the two tasks. In the following, all replication checks, hypotheses, and exploratory tests are listed (related ones follow one another).

First, we tested the following expectations as replication checks (see Table 1):

1. a SNARC effect in both MC and PJ with the standard analysis of a continuous linear regression (this positive control was used as a basis for all further analyses, i.e., finding a SNARC effect with the standard analysis in both tasks is a prerequisite for testing the three hypotheses in this study);
2. a MARC effect in PJ, but not in MC, because the activation of parity seems not to be automatic if only a semantic number feature other than parity is task-relevant;
3. shorter RT in MC than in PJ, because processing magnitude is more straightforward and automatized than processing parity;
4. a Numerical Distance Effect in MC, which is typically found.

To summarize our three hypotheses derived above, we expected (see Table 2):

1. a categorical MC-SNARC, i.e., a better fit of a categorical than continuous MC-SNARC model
2. a continuous PJ-SNARC, i.e., a better fit of a continuous than categorical PJ-SNARC model;
3. (a) a Numerical Size Effect in both tasks, (b) which is stronger in MC than in PJ, because processing magnitude is task-relevant in MC but task-irrelevant in PJ.

Moreover, we explored whether the following observations can be made (see Table 3):

1. task-order effects (Block Orders 1 and 2 versus Block Orders 3 and 4) on both (a) the MC-SNARC and (b) the PJ-SNARC;
2. compatibility-order effects on (a) the MC-SNARC (SNARC slopes in Block Orders 1 and 3 versus Block Orders 2 and 4) and on (b) the PJ-MARC (MARC slopes in Block Orders 1 and 3 versus Block Orders 2 and 4);
3. a good model fit when including both continuous and categorical magnitude predictors for (a) the MC-SNARC or for (b) the PJ-SNARC, indicating a mixed shape of the SNARC effect (see Panel C in Figure 1);
4. number 5 as the most likely boundary between "small" and "large" numbers for most individuals both in MC and PJ; and, relatedly, a categorical MC-SNARC (Hypothesis 1) and a continuous PJ-SNARC (Hypothesis 2), even when comparing the favored categorical model for each participant with the continuous model;
5. a shape difference of (a) the MC-SNARC or for (b) the PJ-SNARC between earlier and later phases within each task;
6. a stronger MC-SNARC than PJ-SNARC when the continuous magnitude predictor is used;
7. a positive correlation between categorical MC-SNARC slopes and continuous PJ-SNARC slopes.

We collected data for MC and PJ with the numbers from 1 to 9 (excluding 5) in a within-subjects design using the typical bimanual response setup. Participants were assigned to one of four block orders, and 30 repetitions provided reliable estimates per experimental cell (number magnitude * response side * task; see Cipora & Wood, 2017). Conducting this study online offered the possibility to test much larger samples than in most previous studies and thus reach high statistical power (Reips, 2000, 2002). The SNARC ef-

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

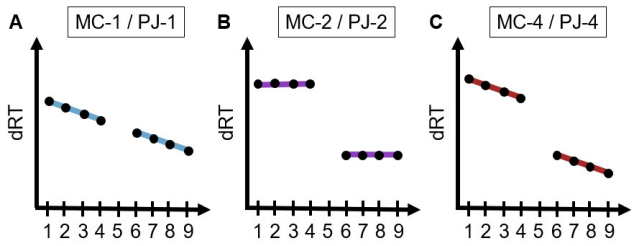


Figure 1. Different shapes of the SNARC effect

Note. In Panel A, the SNARC effect is reflected by a linear regression line with a negative slope, that is, dRT are best predicted by continuous magnitude (models MC-1 and PJ-1). In Panel B, the SNARC effect is reflected by a step-wise function, that is, dRT are best predicted by categorical magnitude (models MC-2 and PJ-2), where all numbers smaller than 5 have the same dRT and all numbers larger than 5 have the same dRT. Panel C shows a mixed shape of the SNARC effect, where both continuous and categorical magnitude predict dRT (models MC-4 and PJ-4). The typically observed MC-SNARC appears as shown in Panel B (Hypothesis 1), and the typically observed PJ-SNARC appears as shown in Panel A (Hypothesis 2). For an overview of all mentioned models, see Table 5. The figure was created using Microsoft PowerPoint.

fect and its various forms have been successfully replicated in online settings (Bulut et al., 2025; Cipora, Soltanlou, et al., 2019; Gökaydin et al., 2018; Hohol et al., 2022; Koch et al., 2023; Roth, Caffier, Reips, Cipora, et al., 2025; Roth, Caffier, Reips, Nuerk, et al., 2025; Roth, Huber, et al., 2025). The measurement in the online setup showed reliability and a similar magnitude compared to the SNARC effect that is typically observed in lab studies. Further, it seems to be valid regarding correlations with mean RT and standard deviations of RT. We calculated Bayes Factors (BF_{10}) to quantify evidence both for differences between MC and PJ as well as for the relationship between the SNARC effects in the two tasks, and lack of such differences or such a relationship. This way, we could shed more light on different SNARC effects specifically in terms of their shape in the two most popular and widely used tasks.

Method

The ethics committee of the University of Tübingen's Department of Psychology approved this study. After peer review and prior to data collection, the Stage-1 Registered Report (https://doi.org/10.17605/OSF.IO/968AD) received in-principle acceptance by the Peer-Community In on March 3rd, 2025. The final Stage-2 Registered Report was recommended by the Peer-Community In on June 12th, 2026 (https://doi.org/10.24072/pci.rr.101213).

Sample Size Considerations

The "Sequential Bayes Factor with maximal n" (SBF+maxN) approach described by Schönbrodt and Wagenmakers (2018)[6] was applied to make our data collection efficient. This means that we ran the data analysis with a total of 500 participants in the first round and recruited further participants sequentially in steps of 50 until the stopping criterion or maximal sample size was reached. Our stopping criterion was moderate evidence regarding all three hypotheses, so that the data either provides evidence in favor of (BF_{10} > 3) or against (BF_{10} < 1/3) each of them.

For the SBF+maxN approach, we defined a maximal (and a minimal) sample size. Thus, we determined the sample size that was necessary to detect evidence for a true underlying effect or for true absence of an effect with a high probability (similar to power analyses in the frequentist framework). This was done for each of the three hypotheses and the largest required sample size was chosen as maximal sample size for the SBF+maxN approach. Our main aim of the current study was to determine the shape of the SNARC effect in each of the two most common tasks. For this, we compared the fit of a continuous and a categorical statistical model in MC and PJ separately (to test Hypotheses 1 and 2). We therefore chose the effect size of interest (ESOI) in a standardized unit, namely Cohen's d = 0.2 (although this is not ideal for power simulations, see Correll et al., 2020). The sample size considerations were based on this ESOI for all three hypotheses, because smaller effect sizes are not practically meaningful. Specifically, d = 0.2 reflects a small effect and corresponds to around 1% of explained variance (calculated using the conversion formula assuming equal-sized groups in Table 2 in Ruscio, 2008). Regarding the detection of a SNARC effect while assuming similar standard deviations as reported in the literature, d = 0.2 corresponds to slopes of -4 ms in the continuous MC-SNARC (with SD = 20 ms), -10 ms in the categorical MC-SNARC (with SD = 50 ms), and -2 ms in the continuous PJ-SNARC (with SD = 10 ms) in their measured unit (i.e., increase of right-hand advantage per continuous magnitude or categorically for large compared to small numbers in milliseconds). These SNARC slopes are of a typically observed or even smaller size.

Analogously to statistical power simulations in the frequentist framework, we randomly drew 5000 samples from a distribution around the ESOI (d = 0.2) and simulated the probability to obtain at least moderate evidence (i.e., BF_{10} > 3) for that effect size by looking at the proportion of Bayesian tests revealing at least moderate evidence for the alternative hypothesis (for a similar approach, see Kelter, 2021; Roth, Caffier, Reips, Nuerk, et al., 2025). Similarly, we randomly drew 5000 samples from a distribution with the respective SD around a truly absent effect (d = 0) and simulated the probability to obtain at least moderate evidence for the null hypothesis (see Kelter, 2021). We thereby determined the sample size that was required for a probability of .90 to obtain moderate evidence for and against the three hypotheses with two-sided Bayesian paired or one-sample t-tests. The required sample size was set as largest for finding at least moderate Bayesian evidence for a true underlying effect of d = 0.2 with a probability of .90 in a two-sided Bayesian independent-samples t-test (n = 2 * 850 = 1700).

6 Note that, apart from a maximum sample size, a minimum sample size needs to be defined as well, which is why it might be more reasonable to term the approach "sequential Bayes Factor with a minimum and maximum N" as done by Witt (2019).

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

The required sample sizes for finding evidence showing true absence of an effect in an independent-samples *t*-test ($n = 2 * 340 = 680$), for evidence for a true underlying effect in a one-sample or paired *t*-test ($n = 440$), or for evidence showing true absence of an effect in a one-sample or paired *t*-test ($n = 160$) were much smaller. We therefore chose $n = 1700$ as a maximal sample size for the SBF+maxN approach. The exact calculations and results for all tests can be found here: https://osf.io/4wpv6/.

Participants

We sequentially recruited adults aged between 18 and 40 years via the recruiting platform Prolific (note that participants' age is self-reported during signup, which can be an issue in recruitment for online experiments, see Reips, 2021). As the study was conducted in English, participation was only possible for native English speakers (as per Prolific's screening based on self-reports). Complete participation was compensated with £5 (Prolific users receive their payment in British pounds), and incomplete participation was compensated partially.

Design and Experimental Task

The present study followed a 2 (task: MC vs. PJ) * 2 (compatibility: incompatible vs. compatible) within-subjects design, resulting in four experimental blocks per participant. Participants were randomly assigned to one of four block orders that resulted from the constraint that both compatibility versions of the same task should follow directly after each other to avoid mixing up instructions (see Figure 2). In Block Orders 1 and 2, participants completed MC in the first and PJ in the second half of the experiment, while the task order was reversed in Block Orders 3 and 4. Within each task, participants were assigned to the SNARC-/MARC-incompatible block first and to the SNARC-/MARC-compatible block second in Block Orders 1 and 3, while the compatibility order was reversed in Block Orders 2 and 4. Given the number of assessed trials (see below), each of the two tasks took approximately 15 minutes, resulting in a median time needed for full participation (including both tasks and some demographic and other questions like the seriousness check) of 33 minutes and 58 seconds.

A binary response-key setup was employed, requiring participants to respond as quickly and accurately as possible using a left or right key (defaults: D or K – can be adjusted individually by participants due to large technical variance on the Internet; Reips, 2000, 2021) depending on whether the number (1, 2, 3, 4, 6, 7, 8, 9) presented on the screen was smaller or larger than 5 (MC) or whether it was odd or even (PJ). Thirty repetitions per experimental cell lead to 240 SNARC-incompatible and 240 SNARC-compatible trials in MC, as well as 240 MARC-incompatible and 240 MARC-compatible trials in PJ per participant. Participants had to take a break of a minimum of 30 seconds between blocks. The order of stimuli within each block was randomized, but each stimulus was presented for the 1st throughout 15th time before any were presented for the 16th throughout 30th time. Each block was divided into two subblocks indistinguishable to the participant, in which each stimulus was presented 15 times (note that this was necessary to investigate Exploratory 5). Each trial started with a square (extended ASCII 254, size 72px), serving as the eye fixation point (300 ms), presented in the center of the screen. Then the number (Open Sans font, size 72px) replaced the square and remained on the screen until a response was given. A blank gray screen (500 ms) concluded the trial. Stimuli as well as fixation squares were presented in black color (0, 0, 0 in RGB notation), while the background remained gray (150, 150, 150 in RGB notation) throughout the experiment. A practice session with 16 trials preceded each block, in which each number was presented twice. Accuracy feedback appeared during practice sessions only.

Procedure

At the beginning of the experiment, a seriousness check (e.g., Reips, 2009) was applied, i.e., participants were asked whether they intended to participate seriously. Participants were asked to take part only if they gave their informed consent, if they used a computer (participation from mobile devices was not possible because a keyboard was required), and if they were between 18 and 40 years old. Then, participants were asked to provide basic demographic data, namely age, gender, first native language, and handedness. They were also asked about their finger counting habits, i.e., which hand they usually start counting with (*right*, *left*, *do not know* or *no preferred hand*) and how stable their preference is (*always*, *mostly*, *slightly more often than with the other hand*, *do not know* or *no preferred hand*). In each question, participants had the opportunity to click on "I prefer not to answer." Next, participants could choose response keys for the experimental task that are located on the same height and about one hand width apart from each other on their keyboard. The default keys D and K meet these requirements on typical keyboards like QWERTZ, QWERTY, and AZERTY and participants were asked to reject that default and change keys if they used a different keyboard layout. Then, instructions were displayed, and the first block of the experimental task started with its practice trials.

After completion of both experimental tasks, data quality was assessed by asking participants how they would describe their environment during participation (*silent*, *very quiet*, *fairly quiet*, *fairly noisy*, *very noisy,* or *extremely noisy),* whether there were any major distractions during participation (*none*, *one,* or *multiple*), and whether there were any difficulties during participation (*yes* or *no*, text field for comments). Moreover, we asked participants whether they had used their left and right index fingers throughout the experiment, as asked for in the task instructions (*yes*, *partly*, or *no*). Participants were provided with a completion code to be inserted in Prolific and with our research team's contact information.

The experiment was set up with WEXTOR (https://wextor.eu; Reips & Neuhaus, 2002) in its HTML and JavaScript framework and adapted (see demo version at https://exp.wextor.eu/esnarc/task/?demo). Our previous experiments have demonstrated that this software is suitable

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

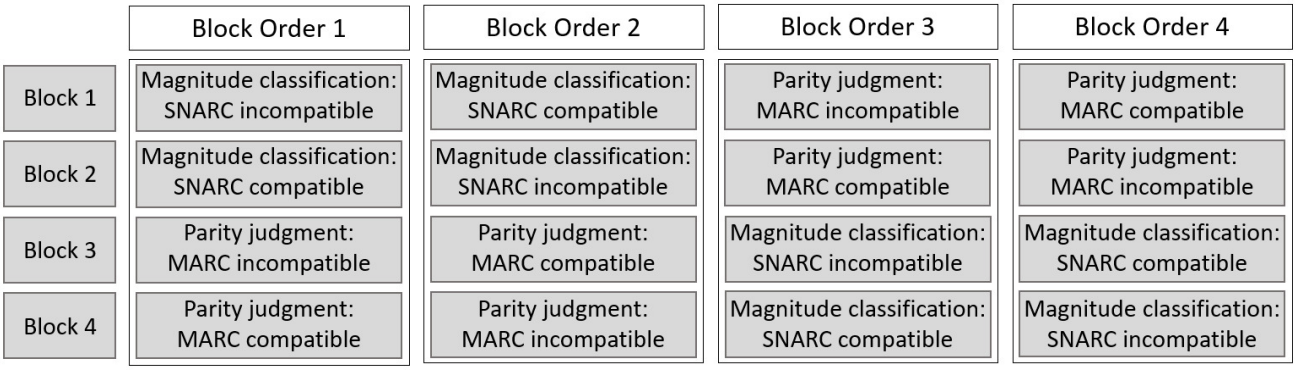


Figure 2. Within-subjects manipulations and resulting block orders counterbalanced between-subjects

Note. We randomly assigned participants to one of the four block orders resulting from the combination of task order and compatibility order in the 2 (task: MC vs. PJ) * 2 (compatibility: incompatible vs. compatible) within-subjects design. The figure was created using Microsoft PowerPoint.

for detecting the SNARC effect in an online setup (Roth, Caffier, Reips, Cipora, et al., 2025; Roth, Caffier, Reips, Nuerk, et al., 2025). To prevent search engine bots (e.g., Googlebot) from submitting data to our experiment, WEXTOR equips the experiment materials with a standardized "noindex, nofollow" meta tag, which prompts search engine bots not to index the experiment pages and also not to visit subsequent pages (see Reips, 2007, p. 379). Further, we restricted participation to devices with a screen width of over 600 pixels. Additionally, to exclude multiple submissions from the same devices, we performed checks based on User-Agents and IP addresses during data evaluation.

Data Preprocessing

All data preprocessing steps and all statistical analyses were performed in the computing software R (R Core Team, 2024). As concerns data preprocessing, we stayed consistent with our previous studies and applied similar inclusion criteria (Roth, Caffier, Reips, Cipora, et al., 2025; Roth, Caffier, Reips, Nuerk, et al., 2025). Only datasets of participants who completed both tasks, who indicated to be between 18 and 40 years old, and who stated their intention to seriously participate were analyzed. Datasets were not included for analyses if participants described their environment as very/extremely noisy, if they reported multiple major distractions, or if participants did not use their left/right index finger for the left/right response key, respectively. Practice trials and incorrectly answered trials were not analyzed. Only trials with RT from 200 to 1500 ms were included in the analysis. Further outliers were removed in an iterative trimming procedure for each participant and task separately, such that only RT that are a maximum 3 SDs above or below the individual mean RT of all remaining trials after these exclusions were considered. Finally, only datasets of participants with at least 75% valid remaining trials per task and without any empty experimental cell (number magnitude * response side * task) were considered.

Data Analysis

We calculated Bayes Factors associated with the corresponding Bayesian *t*-test to obtain evidence for both null and alternative hypotheses (using a default *r*-scale of 0.707 as uninformed prior using Cauchy distribution). A resulting BF_{10} greater than 3 or 10 was treated as moderate or strong evidence for the alternative hypothesis compared to the null hypothesis, respectively, while a resulting BF_{10} smaller than 1/3 or 1/10 was treated as moderate or strong evidence for the null hypothesis compared to the alternative hypothesis, respectively (Dienes, 2021). When using a default *r*-scale of 0.707 Cauchy prior distribution, classifying a BF_{10} larger than 3 as evidence against the null hypothesis is more conservative than rejecting a null hypothesis in the frequentist framework with the typical significance level of $\alpha = .05$ (Wetzels et al., 2011). As explained above, we applied the SBF+maxN approach for sequential data analysis with optional stopping in case of at least moderate evidence for or against each of the three hypotheses.

RT were measured as the time elapsing from the onset of the number presentation on the screen until a response key is pressed (within the limitations that apply in Internet-based research with consumer-grade equipment, see e.g., Garaizar & Reips, 2019). As the dependent variable, we calculated the dRT, which resulted from subtracting the average RT of the left hand from the average RT of the right hand for each number separately per participant and for each task separately.

Several regression models were fit for each participant separately. In these regression models, number magnitude was included as a predictor for dRT to determine the shape of the SNARC effect in each task separately. First, magnitude was included as a continuous predictor, which was equal to the actual stimulus displayed (e.g., 3 for number 3, and 8 for number 8). The resulting regression slopes for *continuous magnitude* represent the advantage of right-hand responses compared to left-hand responses in ms per increase by one in continuous magnitude (i.e., traditional repeated-measures regression in the SNARC effect analysis, as first proposed by Fias et al., 1996). Second, magnitude was contrast-coded as a categorical predictor, using -0.5 for

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

numbers from 1 to 4 and +0.5 for numbers from 6 to 9 (but see Exploratory 4 for a different approach). The resulting regression slopes for *categorical magnitude* represent the advantage of right-hand responses compared to left-hand responses in ms in large compared to small magnitude. Third, for the investigation of the MARC effect, contrast-coded number parity was included as a predictor of dRT, with -0.5 for odd and +0.5 for even numbers (as in Cipora, Soltanlou, et al., 2019). The regression slopes for *parity* represent the advantage of right-hand responses compared to left-hand responses in ms in even compared to odd numbers. An overview of all three predictors (i.e., continuous magnitude, categorical magnitude, and parity), along with their exact coding, can be found in Table 4. For each of the predictors, a more negative coefficient estimate β points towards a stronger SNARC/MARC effect.

Main Data Analysis

An overview of all predictions, analyses, results, and interpretations is given in Table 1 for replication checks, in Table 2 for hypotheses, and in Table 3 for planned exploratory investigations. More details are available in the full Study Design Table (Table B1 in Appendix B) based on a template from the Peer Community In (https://peercommunityin.org), in which all information except outcomes were filled in prior to data collection.

dRT Effects (i.e., SNARC and MARC Effects). For Replication Checks 1 and 2, as well as Hypotheses 1 and 2, we fitted four regression models per participant and per task (for an overview, see Table 5). First, the presence of a SNARC effect in both tasks (Replication Check 1, used as a positive control) was tested in a repeated-measures regression as usually done in SNARC research (see Fias et al., 1996, adapted from Lorch & Myers, 1990). For this, dRT were regressed on continuous magnitude in models MC-1 and PJ-1 for each participant separately. The resulting slopes were tested against zero in a two-sided Bayesian one-sample *t*-test, with Bayesian evidence for a difference from zero indicating a continuous SNARC effect, and negative value of the slope indicating the typical SNARC effect.

Then, the presence of the MARC effect (Replication Check 2) was tested in both tasks with the same repeated-measures regression approach as the SNARC effect. For this, regression models MC-3 and PJ-3 were computed. These models contain a contrast-coded parity predictor and the magnitude predictor with the better fit in the previous test. Because the number-parity predictor and the number-magnitude predictor were orthogonal to each other (i.e., their correlation is zero), both could be concurrently included within one regression model without affecting the respective other parameter estimate. Then, the slopes were tested against zero in a two-sided Bayesian one-sample *t*-test, with evidence for a difference from zero indicating a MARC effect, which is only expected in PJ but not MC.

Next, we investigated whether the MC-SNARC and the PJ-SNARC were continuous or categorical (i.e., which number magnitude predictor fitted the observed dRT better; Hypotheses 1 and 2). For this, besides regressing dRT on continuous magnitude in models MC-1 and PJ-1 as previously described, they were regressed on categorical magnitude (contrast-coded as in Table 4) in MC-2 and PJ-2. Then, we logit-transformed the R^2 for each model for each participant separately to approximate two normal distributions and compare the logit-transformed R^2 between the two models in a two-sided Bayesian paired *t*-test (as Koch et al., 2023, did this in a frequentist approach). In Hypotheses 1 and 2, we predicted a better fit of model MC-2 compared to MC-1 and of PJ-1 compared to PJ-2. Additionally, to support these findings via a Bayesian approach, dRT were regressed on continuous and categorical magnitude for both PJ and MC in four separate Bayesian models, and in each task, a leave-one-out cross validation was performed. An overview of possible SNARC effect shapes and the corresponding regression models tested in the current study can be found in Figure 1.

RT Effects (i.e., Overall Speed, Numerical Size and Distance Effects). For Replication Checks 3 and 4 as well as for Hypotheses 3a and 3b, RT were examined in detail. First, to test whether RT are on average shorter in MC than in PJ (Replication Check 3), we compared mean RT per participant between tasks in a two-sided Bayesian paired *t*-test. Next, the presence of a Numerical Distance Effect in MC (Replication Check 4) and of a Numerical Size Effect in both MC and PJ (Hypotheses 3a and 3b) was investigated with the repeated-measures regression approach (as in Hohol et al., 2020). In MC, RT were regressed on numerical distance (i.e., difference between the number and the criterion number 5) and continuous magnitude (1, 2, 3, 4, 6, 7, 8, or 9) for each participant separately. Because numerical distance and magnitude were orthogonal (i.e., their correlation is zero), both could be concurrently included within one regression model without affecting the respective other parameter estimate. In PJ, RT were only regressed on continuous magnitude for each participant separately. Next, resulting regression slopes were tested against zero in a two-sided Bayesian one-sample *t*-test for each task separately. Evidence for negative slopes for the numerical distance predictor indicated faster responses for larger numerical distance, reflecting the Numerical Distance Effect (Replication Check 4). Evidence for positive slopes for the magnitude predictor indicated slower responses for increasing number magnitude, reflecting the Numerical Size Effect (Hypothesis 3a). Last, we tested whether the magnitude of the Numerical Size Effect was stronger in MC than in PJ (Hypothesis 3b) by comparing resulting slopes between tasks in a two-sided Bayesian paired *t*-test.

Exploratory Data Analysis

Influence of Task and Block Orders. After the analyses for replication checks and hypotheses, we investigated task-order effects on both the MC-SNARC and the PJ-SNARC (Exploratory 1). That is, we tested task-order effects by comparing SNARC slopes in Block Order 1 and 2 (first MC, second PJ) with Block Order 3 and 4 (first PJ, second MC) for each task separately (see Figure 2 for an overview of block orders). The predictor that fitted better in the respective task was used (according to Hypotheses 1 and 2).

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Table 1. Replication checks: Overview of predictions, analyses, results, and interpretations

	Prediction	Analysis	Result	Interpretation
Replication Check 1	A significant SNARC effect is observable in both MC and PJ when determined with the standard analysis of a continuous linear regression.	Regression of dRT on continuous number magnitude (MC-1 and PJ-1) per task and per participant; Two two-sided Bayesian one-sample *t*-tests of SNARC slopes against zero	MC-SNARC: slope = -3.97 ms, *SD* = 13.78 ms, *d* = -0.29, BF_{10} = 1.59 * 10^{9} PJ-SNARC: slope = -3.57 ms, *SD* = 6.36 ms, *d* = -0.56, BF_{10} = 7.85 * 10^{34}	The observed strong evidence for the MC-SNARC and the PJ-SNARC was in line with the hypothesis and previous literature. It laid the groundwork for further hypothesis tests.
Replication Check 2	A MARC effect arises in PJ, but not in MC, because the activation of parity seems not to be automatic when parity is task-irrelevant.	Regression of dRT on contrast-coded number parity (and continuous/ categorical magnitude; MC-3 and PJ-3) per task and per participant; Two two-sided Bayesian one-sample *t*-tests of MARC slopes against zero	MC-MARC: slope = -0.74 ms, *SD* = 23.27 ms, *d* = -0.03, BF_{10} = 0.06 PJ-MARC: slope = -1.36 ms, *SD* = 89.09 ms, *d* = -0.02, BF_{10} = 0.05	The observed strong evidence against a MARC effect was in line with the hypothesis and previous literature for MC, but not for PJ. This discrepancy may result from high variability of the MARC effect between Western individuals.
Replication Check 3	RT are shorter in MC than in PJ because the processing of number magnitude is more automatized than the processing of number parity.	One two-sided Bayesian paired *t*-test to compare mean RT per participant between tasks against zero	$M_{RT\ MC\text{-}PJ}$ = -50.47 ms, *SD* = 48.26, *d* = -1.05, BF_{10} = 2.11 * 10^{96}	The observed strong evidence for faster responses in MC than in PJ was in line with the hypothesis and previous literature.
Replication Check 4	A Numerical Distance Effect arises in MC (i.e., faster responses with increasing numerical distance between the stimulus and the reference number 5).	Regression of RT on numerical distance (and continuous magnitude) for each participant separately; One two-sided Bayesian one-sample *t*-test of numerical-distance slopes against zero	slope = -10.85 ms, *SD* = 7.15 ms, *d* = -1.52, BF_{10} = 4.70 * 10^{156}	The observed strong evidence for the Numerical Distance Effect was in line with the hypothesis and previous literature.

Note. This is a shortened version of the Study Design Table based on a template from the Peer Community In, providing an overview of Replication Checks. For Hypotheses, see Table 2, and for Exploratories, see Table 3. The full version is provided in Appendix B.

For this, we ran two two-sided Bayesian independent-samples *t*-tests.

Further, we tested compatibility-order effects on the MC-SNARC by comparing SNARC slopes in Block Orders 1 and 3 (first SNARC-incompatible, second SNARC-compatible) with SNARC slopes in Block Orders 2 and 4 (first SNARC-compatible, second SNARC-incompatible; Exploratory 2a). Note that we used the categorical slope here (as it described the MC-SNARC better according to the outcome regarding Hypothesis 1). Similarly, we tested compatibility-order effects on the MARC effect in PJ by comparing MARC slopes in Block Orders 1 and 3 (first MARC-incompatible, second MARC-compatible) with MARC slopes in Block Orders 2 and 4 (first MARC-compatible, second MARC-incompatible; Exploratory 2b). For this, we ran two two-sided Bayesian independent-samples *t*-tests. Evidence for a stronger SNARC/MARC effect in Block Orders 2 and 4 compared to Block Orders 1 and 3 reflected larger compatibility effects when the response-key assignment is first compatible and then incompatible, and vice versa.

Exact Shape of the SNARC Effect. Next, a fourth model including both continuous and categorical magnitude was fitted for both tasks (MC-4 and PJ-4). Both resulting slopes were tested against zero in two-sided Bayesian one-sample *t*-tests, with Bayesian evidence for both slopes being different from zero, indicating a mixed shape of the SNARC effect, as illustrated in Panel C in Figure 1 (Exploratory 3). This would mean that the dRT regression slope is negative within small and within large numbers, while there is a categorical step between numbers 4 and 6 (for an empirical observation of such a pattern, see Figure 2b in Nuerk, Bauer, et al., 2005).

Then, we fitted five categorical models for each participant per task. The models were analogous to MC-2 and PJ-2. They differed regarding the boundary between "small" and "large" numbers (contrast-coded with -0.5 and +0.5, respectively). Specifically, we ran five regression models while classifying 1 to 2 vs. 3 to 9, 1 to 3 vs. 4 to 9, 1 to 4 vs. 6 to 9, 1 to 6 vs. 7 to 9, and 1 to 7 vs. 8 to 9 as "small" and "large" numbers (see Table 4). Note that 1 to 4 vs. 6 to 9 corresponds to the MC-2 or PJ-2 (used to test Hypotheses 1 and 2). For each participant, we determined the most likely underlying categorization by descriptively comparing which of the five models had the best fit to the data in terms of R^2. We present the distribution of favored categorical models across the sample for each task in the Results section. Next, we logit-transformed the R^2 for the favored categorical model and for the continuous model for each participant separately and compared the logit-transformed R^2 between the two models in a two-sided Bayesian paired *t*-test

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Table 2. Hypotheses: Overview of predictions, analyses, results, and interpretations

	Prediction	Analysis	Result	Interpretation
Hypothesis 1	The MC-SNARC is categorical.	Regression of dRT on continuous (MC-1) and categorical magnitude (MC-2); Logit-transformation of R^2 for each model per participant; Comparison of the logit-transformed R^2 in a two-sided Bayesian paired *t*-test Purely Bayesian approach: Leave-one-out cross validation	Continuous model MC-1: R^2 = 0.421, R^2_{logit} = -0.317 Categorical model MC-2: R^2 = 0.470, R^2_{logit} = -0.121 Comparison: BF_{10} = 61046054.00 ELPD difference between models of -551.87 with *SE* = 80.64, indicating that MC-2 outperformed MC-1.	The observed strong evidence for a categorical rather than continuous MC-SNARC was in line with the hypothesis and previous literature.
Hypothesis 2	The PJ-SNARC is continuous.	Regression of dRT on continuous (PJ-1) and categorical magnitude (PJ-2); Logit-transformation of R^2 for each model per participant; Comparison of the logit-transformed R^2 in a two-sided Bayesian paired *t*-test Purely Bayesian approach: Leave-one-out cross validation	Continuous model PJ-1: R^2 = 0.194, R^2_{logit} = -1.426 Categorical model PJ-2: R^2 = 0.164, R^2_{logit} = -1.630 Comparison: BF_{10} = 3744.78 ELPD difference between models of -21.42 with *SE* = 4.94, indicating that PJ-1 outperformed PJ-2.	The observed strong evidence for a continuous rather than categorical PJ-SNARC was in line with the hypothesis and previous literature.
Hypothesis 3	Hypothesis 3a: A Numerical Size Effect arises in both tasks. Hypothesis 3b: The Numerical Size Effect is stronger in MC than in PJ.	Regression of RT on continuous magnitude (and numerical distance) for each participant separately; One two-sided Bayesian one-sample *t*-test of continuous-magnitude slopes against zero; One two-sided Bayesian paired *t*-test between continuous-magnitude slopes	MC: slope = 0.39 ms, *SD* = 3.12 ms, *d* = 0.12, BF_{10} = 4.75 PJ: slope = 1.08 ms, *SD* = 2.67 ms, *d* = 0.40, BF_{10} = $3.17 * 10^{18}$ Comparison: BF_{10} = 249.08 for a difference between tasks	The observed moderate/strong evidence for the Numerical Size Effect was in line with the hypothesis and previous literature. However, in contrast to the hypothesis, the effect was stronger in PJ than in MC.

Note. This is a shortened version of the Study Design Table based on a template from the Peer Community In, providing an overview of Hypotheses. For Replication Checks, see Table 1, and for Exploratories, see Table 3. The full version is provided in Appendix B.

(Exploratory 4). Moreover, we regressed dRT on continuous and categorical magnitude (using the favored categorical predictor for each participant) for both PJ and MC in Bayesian models and performed a leave-one-out cross validation (as for Hypotheses 1 and 2).

As a robustness check, to find out whether the determined boundary between "small" and "large" numbers was reliable for each participant and not only due to random measurement noise, we determined its split-half reliability by splitting valid trials with the odd-even method based on presentation order (i.e., 1st, 3rd, 5th, etc. trial vs. 2nd, 4th, 6th, etc. trial). Subsequently, we again computed five models per participant and per task, separately for each half of all valid trials, and determined which of the five models has the largest R^2 in each half of the experiment. For each half per participant within each task, we coded the boundary of the model with the highest R^2 with the cardinal boundary values 2.5, 3.5, 5, 6.5, or 7.5. We then computed the Pearson product-moment correlation between the favored categorical models across halves within each task (note that coding the favorable model ordinally with 1, 2, 3, 4, and 5 to calculate the Spearman rank correlation would lead to almost the same results, as the correlation between the cardinal and ordinal values is r = .997). Next, we applied the Spearman-Brown correction to each correlation to adjust for task length and descriptively evaluate the correlations in terms of whether the boundaries between "small" and "large" numbers are related between both halves. We also created alluvial plots showing the stability of the boundaries within participants across halves for each task separately. Finally, we computed the split-half reliability for the basic categorical MC-SNARC (using the classification of 1 to 4 vs. 6 to 9 as "small" vs. "large") and continuous PJ-SNARC. These results were taken into account when interpreting the split-half reliability of the favored categorical model (as a part of Exploratory 4). In the literature, a split-half reliability of r = 0.89 has been reported for the categorical MC-SNARC (Hohol et al., 2020, see Supplementary Material 3), and $0.43 \leqslant r \leqslant 0.82$ have been reported for the continuous PJ-SNARC (for an overview, see Table 1 in Cipora, van Dijck, et al., 2019).

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Table 3. Exploratories: Overview of predictions, analyses, results, and interpretations

	Prediction	Analysis	Result	Interpretation
Exploratory 1	There may be task-order effects on the MC-SNARC (Exploratory 1a) and PJ-SNARC (Exploratory 1b).	Two two-sided Bayesian *t*-tests comparing SNARC slopes (continuous/ categorical) in the MC-PJ order with the PJ-MC order for each task separately	MC-SNARC: BF_{10} = 0.09 PJ-SNARC: BF_{10} = 0.15	Evidence was strong or moderate against task-order influences on the MC-SNARC or PJ-SNARC.
Exploratory 2	There may be compatibility-order effects on the MC-SNARC (Exploratory 2a) and PJ-MARC (Exploratory 2b).	Two two-sided Bayesian *t*-tests comparing SNARC slopes (continuous/ categorical) / MARC slopes in the incompatible-compatible order with the compatible-incompatible order for each task separately	MC-SNARC: BF_{10} = 10009.40; slope = -36.40 ms in incompatible-compatible order; slope = -4.45 ms in compatible-incompatible order PJ-MARC: BF_{10} = 0.52; slope = -8.26 ms in incompatible-compatible order; slope = 5.38 ms in compatible-incompatible order	Strong evidence was found for a stronger MC-SNARC when participants completed the incompatible before the compatible block. Evidence regarding compatibility-order influences on the PJ-MARC was inconclusive.
Exploratory 3	The shape of the MC-SNARC (Exploratory 3a) and PJ-SNARC (Exploratory 3b) may be mixed.	Regression of dRT on continuous and categorical magnitude (MC-4 and PJ-4) per task and per participant; Four two-sided Bayesian one-sample *t*-tests of SNARC slopes against zero	MC-SNARC: continuous $BF_{10} = 5.14 * 10^{13}$, categorical BF_{10} = 0.05; MC-4: R^2 = 0.573, R^2_{logit} = 0.294 PJ-SNARC: continuous $BF_{10} = 6.98 * 10^{20}$, categorical BF_{10} = 277.05) PJ-4: R^2 = 0.279, R^2_{logit} = -0.949	When considering both predictors simultaneously, evidence was strong for the continuous but against the categorical one in MC, in contrast with the categorical shape that resulted for Hypothesis 1. Evidence was strong for a mixed shape of the PJ-SNARC. Note that high collinearity between predictors makes these results hardly interpretable.
Exploratory 4	Number 5 is the most likely boundary between "small" and "large" in both tasks. The MC-SNARC remains categorical and the PJ-SNARC remains continuous, even when considering the categorical model favored for each individual.	Regression of dRT on five different categorical magnitude predictors (differing in the boundary between "small" and "large" numbers) per task and per participant; Descriptive comparison of R^2 between the five models (see Table 4); Logit-transformation of R^2 for the favored categorical model and for the continuous model for each participant; Comparison of the logit-transformed R^2 in a two-sided Bayesian paired *t*-test; Purely Bayesian approach: Leave-one-out cross validation	MC: $BF_{10} = 8.90 * 10^{80}$; favored categorical model (maximum for boundary 5 in 49.8% of participants): R^2 = 0.573, R^2_{logit} = 0.293, continuous model: R^2 = 0.421, R^2_{logit} = -0.317 MC: ELPD difference between models of -934.75 with *SE* = 82.06, indicating the favored categorical model outperformed the continuous model PJ: $BF_{10} = 4.74 * 10^{86}$; favored categorical model (maximum for boundary 3.5 in 28.5% participants): R^2 = 0.333, R^2_{logit} = -0.693,	In MC, number 5 was the favored categorical boundary for half the participants, whereas in PJ, no categorical boundary was as clearly favored. The categorical model using the boundary favored for each individual fit the data better than the continuous model in both tasks. Split-half reliabilities for the boundaries between "small" and "large" numbers were low in both tasks. Those for the SNARC effect were high in MC and intermediate in PJ.

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

	Prediction	Analysis	Result	Interpretation
		Determination of the split-half reliability for the boundary as well as for the basic MC-SNARC and PJ-SNARC using the odd-even method	continuous model: R^2 = 0.194, R^2_{logit} = -1.426 PJ: ELPD difference between models of -81.59 with *SE* = 22.37, indicating the favored categorical model outperformed the continuous model Split-half reliabilities (Spearman-Brown corrected): MC boundary: r = 0.06, BF_{10} = 0.13, PJ boundary: r < 0.01, BF_{10} = 0.09 MC-SNARC: r = 0.94, BF_{10} = 2.46 * 10^{209} PJ-SNARC: r = 0.44, BF_{10} = 6.34 * 10^{9}	
Exploratory 5	The shape of the MC-SNARC (Exploratory 5a) or PJ-SNARC (Exploratory 5b) may differ between earlier and later phases within each task.	Recomputation of dRT by only considering the first or last 15 out of 30 repetitions per block within each task per participant; Regression of new dRT on continuous and categorical magnitude (MC-4 and PJ-4) per task and per participant; Eight two-sided Bayesian one-sample *t*-tests of SNARC slopes against zero	MC-SNARC: continuous: $BF_{10\,first15}$ = 33123391, $BF_{10\,last15}$ = 4032107, categorical: $BF_{10\,first15}$ = 0.06, $BF_{10\,last15}$ = 0.05 PJ-SNARC: continuous: $BF_{10\,first15}$ = 3.25 * 10^{13}, $BF_{10\,last15}$ = 2.16 * 10^{9}, categorical: $BF_{10\,first15}$ = 1.89, $BF_{10\,last15}$ = 6.17	In MC, strong evidence for a continuous shape and against a categorical shape was in contrast with results for Hypothesis 1. In PJ, strong evidence for a continuous shape and inconclusive to moderate evidence for a categorical shape was in line with results for Hypothesis 2. The shape of the SNARC effect seems to be similar in earlier and later phases.
Exploratory 6	The MC-SNARC may be stronger than the PJ-SNARC when continuously modeled.	Regression of dRT on continuous magnitude separately per task and per participant; One two-sided Bayesian paired *t*-test between continuous-magnitude slopes	MC: slope = -3.97 ms PJ: slope = -3.57 ms Comparison: BF_{10} = 0.06	We observed strong evidence against different effect strengths of the MC-SNARC and the PJ-SNARC.
Exploratory 7	SNARC slopes from the better fitting models (categorical in MC and continuous in PJ, see Hypotheses 1 and 2) may be positively correlated.	One two-sided Bayesian Pearson correlation test for the categorical MC-SNARC and the continuous PJ-SNARC	r = 0.17, BF_{10} = 423.87 Maximum given reliabilities (Exploratory 4): r_{max} = 0.64	We observed strong evidence for a positive correlation between the categorical MC-SNARC and the continuous PJ-SNARC.

Note. This is a shortened version of the Study Design Table based on a template from the Peer Community In, providing an overview of Exploratories planned at Stage 1 of this Registered Report. For Replication Checks, see Table 1, and for Hypotheses, see Table 2. The full version is provided in Appendix B.

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Table 4. Overview of dRT predictors

Continuous magnitude	1	2	3	4	6	7	8	9
Categorical magnitude: Boundary 5	-0.5	-0.5	-0.5	-0.5	+0.5	+0.5	+0.5	+0.5
Boundary 2.5 (Exploratory 4)	-0.5	-0.5	+0.5	+0.5	+0.5	+0.5	+0.5	+0.5
Boundary 3.5 (Exploratory 4)	-0.5	-0.5	-0.5	+0.5	+0.5	+0.5	+0.5	+0.5
Boundary 6.5 (Exploratory 4)	-0.5	-0.5	-0.5	-0.5	-0.5	+0.5	+0.5	+0.5
Boundary 7.5 (Exploratory 4)	-0.5	-0.5	-0.5	-0.5	-0.5	-0.5	+0.5	+0.5
Parity	-0.5	+0.5	-0.5	+0.5	+0.5	-0.5	+0.5	-0.5

Note. This table gives an overview of the dRT predictors that were used in the regression models summarized in Table 5. Continuous magnitude is equal to the actual presented stimulus. Categorical magnitude was contrast-coded with -0.5 for smaller vs. +0.5 for larger numbers (Boundary 5, i.e., a categorization into numbers 1 to 4 vs. 6 to 9, is used for most analyses, whereas different boundaries are used in Exploratory 4). Number parity was always contrast-coded with -0.5 for odd and +0.5 for even numbers.

Table 5. Overview of regression models that were fitted for each participant

Magnitude classification	
MC-1	$dRT \sim \beta_0 + \beta_1 * magnitude_{continuous}$
MC-2	$dRT \sim \beta_0 + \beta_1 * magnitude_{categorical}$
MC-3[a]	$dRT \sim \beta_0 + \beta_1 * magnitude_{continuous/categorical} + \beta_2 *$ parity
MC-4	$dRT \sim \beta_0 + \beta_1 * magnitude_{continuous} + \beta_2 *$ $magnitude_{categorical}$
Parity judgment	
PJ-1	$dRT \sim \beta_0 + \beta_1 * magnitude_{continuous}$
PJ-2	$dRT \sim \beta_0 + \beta_1 * magnitude_{categorical}$
PJ-3[b]	$dRT \sim \beta_0 + \beta_1 * magnitude_{continuous/categorical} + \beta_2 *$ parity
PJ-4	$dRT \sim \beta_0 + \beta_1 * magnitude_{continuous} + \beta_2 *$ $magnitude_{categorical}$

Note. Four regression models were fitted for each participant separately for MC (MC-1, MC-2, MC-3, and MC-4) and PJ (PJ-1, PJ-2, PJ-3, and PJ-4). The predictors used in the models are specified in Table 4. In each model, β_1 and β_2 are the coefficients of the respective predictors for number magnitude or number parity. β_0 denotes the model intercept. [a] For MC-3, the better magnitude predictor from MC-1 and MC-2 was used. [b] For PJ-3, the better magnitude predictor from PJ-1 and PJ-2 was used.

Time Course of the SNARC Effect. Moreover, we explored whether the shape of the SNARC effect differed between earlier and later phases within each task (Exploratory 5). Importantly, it was not possible to determine the SNARC effect in the first or second block of each task separately, because both blocks are needed in order to calculate the differences between left- and right-hand responses. Therefore, we computed the models MC-4 and PJ-4 and tested the resulting slopes for both a continuous predictor and a categorical predictor against zero in two-sided Bayesian one-sample *t*-tests, but instead of considering all 30 repetitions per block, we only considered the first or second halves of both blocks within each task (i.e., first or second 15 repetitions of each number in one and in the other response-to-key assignment). This way, we could investigate whether early trials in each response-to-key assignment lead to a different SNARC shape than late trials.

Relations Between the MC-SNARC and the PJ-SNARC. Further, we tested whether the MC-SNARC is stronger than the PJ-SNARC by comparing the slopes for continuous magnitude predictors resulting from models MC-1 and PJ-1 in a two-sided Bayesian paired *t*-test (Exploratory 6). Lastly, we calculated Pearson's correlation between the MC-SNARC and PJ-SNARC slopes (Exploratory 7). For this, the predictor leading to a better model fit per task across all participants was used (categorical in MC according to Hypothesis 1 and linear in PJ according to Hypothesis 2). We ran a two-sided Bayesian Pearson correlation test to see whether the spatial mapping of number magnitude within participants was similar in both tasks.

Data Quality and Positive Controls

To control the data quality in our study, we implemented a seriousness check (Reips, 2009) as well as a self-assessment of noise, distractions, and other difficulties. To make sure that we only analyzed trials that reflected mental processes in correctly executed MC or PJ, we only included correctly answered trials, trimmed RT, and only included data from participants with a minimum of 75% remaining valid trials (as described in the data preprocessing pipeline). Moreover, the test of the MC-SNARC and PJ-SNARC analyzed with the traditional linear regression (Replication Check 1) served as positive control. Importantly, we considered this positive control as a prerequisite for all further analyses and only proceeded with testing the other hypotheses, because we could find at least moderate Bayesian evidence for both the continuous MC-SNARC and the continuous PJ-SNARC at the group level. Finally, in additional replication checks, we aimed to replicate results from previous studies to validate our investigation.

Possible Limitations and Unexpected Outcomes

Importantly, including both continuous and categorical magnitude within one single regression model (as in MC-4 and PJ-4) is problematic because of collinearity between these two predictors (i.e., r = .913). However, we still decided to compute one such regression model for each task because the true shape of the SNARC effect might be determined by both continuous and categorical number magnitude simultaneously.

In the present study, we tested the two most frequently used versions of MC and PJ (i.e., with symbolic single-digit numbers) in a sample in which the SNARC effect is not

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

controversial (i.e., Western culture with left-to-right reading and writing direction). Future studies will show whether our results hold true for different types of stimuli and for different samples.

Results

The data collection and the data analyses were conducted as described in the peer-reviewed Stage-1 Registered Report accepted by the Peer Community In. Additional non-preregistered analyses can be found in Appendix C. All R scripts for data preprocessing and analysis as well as all anonymized datasets can be found at https://osf.io/g48s2/. The following R packages were used: *BayesFactor* by Morey et al. (2024), *bayestestR* by Makowski et al. (2019), *brms* by Buerkner (2017), *data.table* by Barrett et al. (2024), *dplyr* by Wickham, François et al. (2023), *ggalluvial* by Brunson & Read (2023), *ggplot2* by Wickham et al. (2024), *ggpubr* by Kassambara (2025), *loo* by Vehtari et al. (2017), *scales* by Wickham et al. (2025), and *tidyr* by Wickham, Vaughan, et al. (2023). Observed outcomes were added to Tables 1, 2, and 3, as well as to the Study Design Table (see rightmost column of Table B1 in Appendix B) that was filled in prior to data collection.

Data Preprocessing

Our final sample recruited in line with the SBF+maxN approach consisted of 701 participants who completed both tasks. Out of these participants, none were reported to be under 18 and three were reported to be over 40. Two indicated they only wanted to have a look at the experiment but not seriously participate. Further, one participant reported an extremely noisy environment and two reported multiple major distractions. Thirty-one participants did not use their left/right index finger for the left/right response key consistently as described in the instructions. In total, the data of 39 participants were therefore excluded from the analysis in a first preprocessing step. Next, while ignoring practice trials, exclusion criteria were applied at the trial level for each task separately:

- MC: 5.05% of the trials were excluded due to incorrect responses, 0.30% due to responses faster than 200 ms, 2.23% due to responses slower than 1500 ms, and 7.09% in the iterative trimming procedure.
- PJ: 5.95% of the trials were excluded due to incorrect responses, 0.20% due to responses faster than 200 ms, 2.96% due to responses slower than 1500 ms, and 6.56% in the iterative trimming procedure.

Finally, after these exclusions at the trial level, 13 participants had at least one empty experimental cell, and 51 participants had less than 75% valid trials remaining (30 in MC and 33 in PJ, of which 12 overlappingly in both tasks). Their data were thus excluded from analyses, and data from 611 participants remained for analysis. Originally, we had planned to apply dropout analysis (Reips et al., 2025) to detect possible motivational confounding (Reips, 2000, 2021), but given there were only few dropouts, such an analysis was not necessary.

Descriptive self-reported information about the participants is summarized in Table 6. Most of them used the default response keys, but 38 replaced D and 41 replaced K (among whom 35 replaced both keys). Parameter estimates for the four main regression models (see Table 5) can be found in Table 7.

Data Analysis

Main Data Analysis

The average RT for all included trials were 526 ms (*SD* = 100 ms) in MC and 577 ms (*SD* = 100 ms) in PJ. Figure 3 illustrates the mean RT including their standard error (SE) per stimulus and per response in each task. Information about accuracy rates and potential speed-accuracy trade-offs is provided in Exploratory 19 in Appendix C.

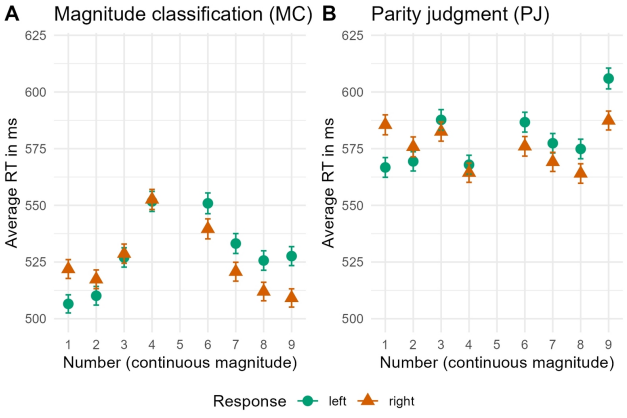


Figure 3. Mean RT (error bars represent ± 1 SE) in MC (Panel A) and PJ (Panel B) per digit and per response side for trials included in the data analysis

Replication Check 1: Standard SNARC Effect. Bayesian one-sample *t*-tests of the continuous magnitude slopes against zero revealed strong evidence for an MC-SNARC (slope = -3.97 ms, *SD* = 13.78 ms, *d* = -0.29, BF_{10} = 1.59 * 10^9), and strong evidence for a PJ-SNARC (slope = -3.57 ms, *SD* = 6.36 ms, *d* = -0.56, BF_{10} = 7.85 * 10^{34}). The positive control was thus successful, which is in line with our expectation and fulfills the prerequisite for testing all further hypotheses. The MC-SNARC is plotted in Figure 4, and the PJ-SNARC is plotted in Figure 5.

Replication Check 2: Standard MARC Effect. Bayesian one-sample *t*-tests of the contrast-coded parity slopes against zero revealed strong evidence against an MC-MARC (slope = -0.74 ms, *SD* = 23.27 ms, *d* = -0.03, BF_{10} = 0.06), as well as against a PJ-MARC (slope = -1.36 ms, *SD* = 89.09 ms, *d* = -0.02, BF_{10} = 0.05). Note that this is at odds with several studies demonstrating the MARC effect in PJ (i.e., no replication).

Replication Check 3: RT Differences Between Tasks. A two-sided Bayesian paired *t*-test comparing mean RT averaged across all trials within each participant revealed strong evidence for shorter RT in MC than in PJ ($M_{RT\ MC}$ = 526.07 ms, $M_{RT\ PJ}$ = 576.54 ms, $M_{RT\ MC\text{-}PJ}$ = -50.47 ms, *d* =

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Table 6. Descriptive self-reported information about the sample (N = 611 after exclusions)

Demographic item	Answer	Count (Proportion)
Age in years	Mean (SD)	30.0 (5.62)
Gender	Woman	297 (48.6%)
	Man	311 (50.9%)
	Other	3 (0.5%)
Native language	English	610 (99.8%)
	Others	1 (0.2%)
Handedness	Right-handed	532 (87.1%)
	Left-handed	56 (9.2%)
	Ambidextrous	23 (3.8%)
Finger counting habits	Right-Starters	339 (55.5%)
	Left-Starters	220 (36.0%)
	Does not know or no preference	52 (8.5%)
Finger counting stability	Always	276 (45.2%)
	Mostly	241 (39.4%)
	Slightly more often	38 (6.2%)
	Does not know or no preference	2 (0.3%)
	No answer	54 (8.8%)

Table 7. Summary of the predictors in both tasks (i.e., mean and SD estimates) using models in Table 5

MC			PJ		
Model	**Predictor**	**Mean slope (SD)**	**Model**	**Predictor**	**Mean slope (SD)**
MC-1	β_1	-3.97 (13.78)	PJ-1	β_1	-3.57 (6.36)
MC-2	β_1	-20.25 (81.80)	PJ-2	β_1	-16.19 (32.16)
MC-3	β_1 β_2	-13.77 (76.37) -0.74 (23.27)	PJ-3	β_1 β_2	-7.58 (22.20) -1.36 (89.09)
MC-4	β_1 β_2	-3.60 (10.35) -2.26 (95.71)	PJ-4	β_1 β_2	-5.24 (12.24) 9.98 (58.55)

-1.05, BF_{10} = 2.11 * 10^{96}). This result is in line with our expectation and previously reported results (see Figure 3).

Replication Check 4: Numerical Distance Effect. A two-sided Bayesian one-sample *t*-test of the numerical distance predictor for RT against zero provided strong evidence for a Numerical Distance Effect in MC (slope = -10.85 ms, *SD* = 7.15 ms, *d* = -1.52, BF_{10} = 4.70 * 10^{156}). This replicates faster responses in MC for larger numerical distances (see Panel A in Figure 3).

Hypothesis 1: Shape of MC-SNARC. A two-sided Bayesian paired *t*-test revealed strong evidence for a better fit of the categorical model MC-2 (R^2 = 0.470, R^2_{logit} = -0.121) than of the continuous model MC-1 (R^2 = 0.421, R^2_{logit} = -0.317) with BF_{10} = 61046054.00. In line with our hypothesis, this result indicates a rather categorical than continuous shape of the MC-SNARC. This was further supported by an ELPD difference (MC-1 vs. MC-2) of -551.87 (*SE* = 80.64) obtained in the leave-one-out cross validation, indicating that MC-2 outperformed MC-1.

Hypothesis 2: Shape of PJ-SNARC. A two-sided Bayesian paired *t*-test revealed strong evidence for a better fit of the continuous model PJ-1 (R^2 = 0.194, R^2_{logit} = -1.426) than of the categorical model PJ-2 (R^2 = 0.164, R^2_{logit} = -1.630) with BF_{10} = 3744.78. In line with our hypothesis, this result indicates a rather continuous than categorical shape of the PJ-SNARC. This was further supported by an ELPD difference (PJ-1 vs. PJ-2) of -21.42 (*SE* = 4.94) obtained in the leave-one-out cross validation, indicating that PJ-1 outperformed PJ-2.

Hypotheses 3a and 3b: Numerical Size Effect. Two-sided Bayesian one-sample *t*-tests of the continuous magnitude predictor for RT against zero provided moderate evidence for a Numerical Size Effect in MC (slope = 0.39 ms, *SD* = 3.12 ms, *d* = 0.12, BF_{10} = 4.75) and strong evidence for a Numerical Size Effect in PJ (slope = 1.08 ms, *SD* = 2.67 ms, *d* = 0.40, BF_{10} = 3.17 * 10^{18}). This is in line with faster responses for smaller than larger number magnitudes in both tasks, as predicted in Hypothesis 3a and as previously reported in the literature. However, a two-sided Bayesian paired *t*-test revealed strong evidence for a stronger Numerical Size Effect in PJ than in MC (BF_{10} = 249.08), which is the opposite of what was expected in Hypothesis 3b.

Exploratory Data Analysis

Exploratory 1: Task-order Effects on the SNARC Effect. Two-sided Bayesian independent-samples *t*-tests revealed strong evidence against task-order effects both on the categorical MC-SNARC (BF_{10} = 0.09; slope = -20.74 ms in Block Orders 1 and 2, and slope = -19.79 ms Block Orders 3 and 4), and on the continuous PJ-SNARC (BF_{10} = 0.15;

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

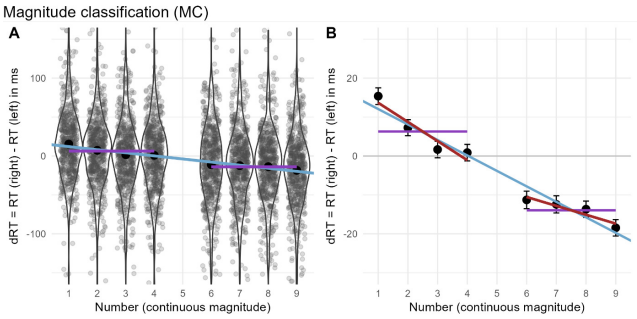


Figure 4. The SNARC effect in MC

Note. Panels A and B both illustrate the same data from MC, with large black dots representing mean dRT per number magnitude across all participants. The blue lines with the negative slope correspond to the continuous model MC-1, the purple horizontal lines correspond to the categorical model MC-2, and the brown lines in Panel B illustrate negative slopes for continuous regressions calculated separately for numbers from 1 to 4 and from 6 to 9 (see Exploratory 17). In Panel A, smaller gray dots summarized by violins represent mean dRT per number magnitude per participant separately. In Panel B, error bars represent ± 1 SE. Please note that the y-axis scaling differs between Panels A and B.

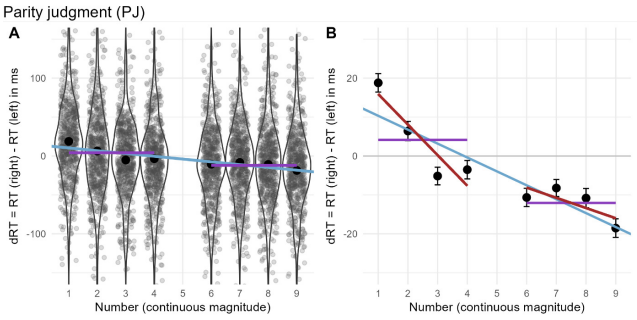


Figure 5. The SNARC effect in PJ

Note. Panels A and B both illustrate the same data from PJ, with large black dots representing mean dRT per number magnitude across all participants. The blue lines with the negative slope correspond to the continuous model PJ-1, the purple horizontal lines correspond to the categorical model PJ-2, and the brown lines in Panel B illustrate negative slopes for continuous regressions calculated separately for numbers from 1 to 4 and from 6 to 9 (see Exploratory 17). In Panel A, smaller gray dots summarized by violins represent mean dRT per number magnitude per participant separately. In Panel B, error bars represent ± 1 SE. Please note that the y-axis scaling differs between Panels A and B.

slope = -3.29 ms in Block Orders 1 and 2, and slope = -3.83 ms Block Orders 3 and 4).

Exploratory 2: Compatibility-order Effects on the SNARC Effect. Two-sided Bayesian independent-samples *t*-tests revealed strong evidence for a compatibility-order effect on the categorical MC-SNARC (BF_{10} = 10009.40; slope = -4.45 ms in Block Orders 2 and 4, starting with SNARC-compatible, and slope = -36.40 ms in Block Orders 1 and 3, starting with SNARC-incompatible). Evidence regarding a compatibility-order effect on the continuous PJ-MARC was inconclusive (BF_{10} = 0.52; slope = 5.38 ms in Block Orders 2 and 4, starting with MARC-compatible, and slope = -8.26 ms in Block Orders 1 and 3, starting with MARC-incompatible).

Exploratory 3: Mixed SNARC Shape Within Joint Models. The continuous and categorical SNARC slopes resulting from a joint model using both predictors (MC-4 or PJ-4) were tested against zero separately in two-sided Bayesian one-sample *t*-tests. Evidence was strong for a continuous shape ($BF_{10} = 5.14 * 10^{13}$) and against a categorical shape (BF_{10} = 0.05) of the MC-SNARC. Evidence was strong for both a continuous ($BF_{10} = 6.98 * 10^{20}$) and a categorical shape (BF_{10} = 277.05) of the PJ-SNARC. Naturally, the fits were higher for MC-4 (R^2 = 0.573, R^2_{logit} = 0.294) and PJ-4 (R^2 = 0.279, R^2_{logit} = -0.949) than for MC-2 and PJ-1, which only included the more suitable magnitude predictor but not both. These results are not in line with the categorical MC-SNARC observed in Hypothesis 1 and point towards a mixed shape of the PJ-SNARC (see Panel C in Figure 1). This might be due to high collinearity between predictors reducing the explained variance attributable to each separately, complicating interpretation. Additionally, the model is limited by the number of data points (i.e., eight stimuli) being small relative to the number of parameters estimated (i.e., two magnitude predictors). To overcome these limitations, we additionally ran Exploratory 17 (see Appendix C). The reasoning was that, on the one hand, the MC-SNARC seems to follow a step-wise pattern more than a linear one (Hypothesis 1). On the other hand, as visible in Figure 4 and as revealed by a similarly high fit of the continuous model, there might still be a linear trend within small and large numbers. We therefore tested continuous regression slopes within small and large numbers separately in Exploratory 17.

Exploratory 4: Categorization into "Small" and "Large" Numbers. The distribution of the most likely underlying categorical boundaries between "small" and "large" numbers obtained by comparing five models with different boundaries (2.5, 3.5, 5, 6.5, or 7.5; see Table 4) per participant in each task is summarized in Table 8. Naturally, in MC, stronger evidence was found for the fit of the favored categorical (R^2 = 0.573, R^2_{logit} = 0.293) rather than continuous model MC-1 (R^2 = 0.421, R^2_{logit} = -0.317) with $BF_{10} = 8.90 * 10^{80}$. In line with this, the ELPD difference obtained in the leave-one-out cross validation was -934.75 (*SE* = 82.06), suggesting a better fit for the favored categorical model than for the continuous MC-1 model. Interestingly, in PJ, strong evidence was found for a better fit of the favored categorical (R^2 = 0.333, R^2_{logit} = -0.693) than continuous model PJ-1 (R^2 = 0.194, R^2_{logit} = -1.426) with $BF_{10} = 4.74 * 10^{86}$ as well, although the results for Hypothesis 2 revealed a better fit of the continuous model when compared with the standard categorical model. In line with this, the ELPD difference obtained in the leave-one-out cross validation was -81.59 (*SE* = 22.37), also indicating a better fit of the favored categorical model. Evidence regarding the split-half reliability of the favored categorical boundaries was moderately against a difference from zero in MC (Spearman-Brown corrected estimate *r* = 0.06; BF_{10} = 0.13) and strongly against a difference from zero in PJ (Spearman-Brown corrected estimate *r* < 0.01; BF_{10} = 0.09), as calculated with the odd-even method. Thus, categorical boundaries between "small" and "large" seemed rather unreliable, as illustrated in Figure 6. In contrast, evidence for the split-half reliability of the SNARC effect in each task to differ from zero was strong: The odd-even method revealed *r* = 0.89 for the standard categorical MC-SNARC (Spearman-Brown corrected estimate *r* = 0.94; $BF_{10} = 2.46 * 10^{209}$) and *r*

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Table 8. Number and proportion of participants for each favored categorical boundary between "small" and "large" (2.5, 3.5, 5, 6.5, or 7.5) in each task (Exploratory 4).

	Number of participants (proportion)	
Favored categorical model	MC	PJ
Boundary 2.5	65 (10.6 %)	111 (18.2 %)
Boundary 3.5	90 (14.7 %)	174 (28.5 %)
Boundary 5	304 (49.8 %)	72 (11.8 %)
Boundary 6.5	96 (15.7 %)	164 (26.8 %)
Boundary 7.5	56 (9.2 %)	90 (14.7 %)

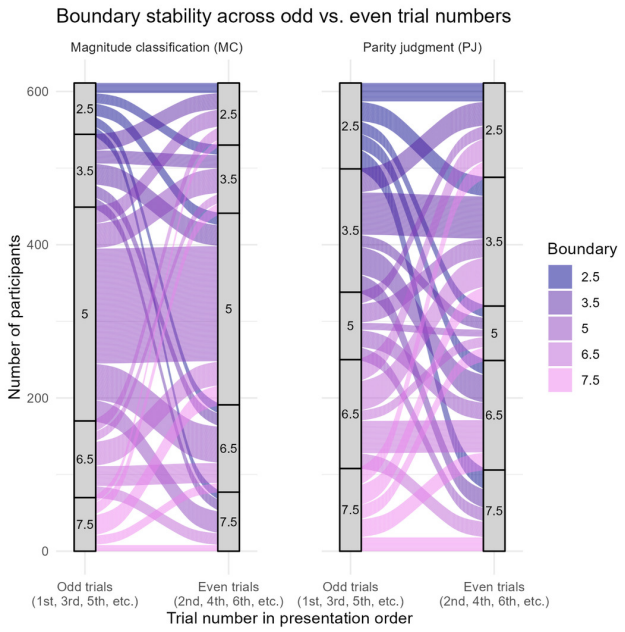


Figure 6. Alluvial plots illustrating the split-half reliability of categorical boundaries between "small" and "large" numbers per task (Exploratory 4)

Note. These alluvial plots show the proportion of participants for which the favored categorical boundary between "small" and "large" was 2.5, 3.5, 5, 6.5, or 7.5 in each half of each task (left panel: MC, right panel: PJ) when splitting trials based on presentation order (odd: 1st, 3rd, 5th, etc. vs. even: 2nd, 4th, 6th, etc.).

= 0.28 for the continuous PJ-SNARC (Spearman-Brown corrected estimate $r = 0.44$; $BF_{10} = 6.34 * 10^9$). The higher split-half reliability for the MC-SNARC than for the PJ-SNARC is visible in Figure 7.

Exploratory 5: Time Course of the SNARC Effect. Next, we considered the first and last 15 repetitions of each number within each response-to-key assignment by recomputing separate MC-4 and PJ-4 models and testing both predictors in two-sided Bayesian one-sample *t*-tests. Strikingly, evidence was strong for a continuous shape of the MC-SNARC in both the first (BF_{10} = 33123391) and last (BF_{10} = 4032107) 15 repetitions, and strongly against a categorical shape of the MC-SNARC in both the first (BF_{10} = 0.06) and last (BF_{10} = 0.05) 15 repetitions, which is not in line with results for Hypothesis 1. Likewise, evidence was strong for a continuous shape of the PJ-SNARC in both the first ($BF_{10} = 3.25 * 10^{13}$) and last ($BF_{10} = 2.16 * 10^9$) 15 repetitions. Evidence for a categorical shape of the PJ-SNARC was inconclusive in the first (BF_{10} = 1.89) and moderately in favor in the last (BF_{10} = 6.17) 15 repetitions. Note, however, that collinearity between the continuous and categorical predictors was high, so including both in a single model is problematic because each accounts for overlapping proportions of variance. The role of each predictor should be interpreted conditional on the respective other, although the high collinearity makes the distinction between the two unreliable. To sum up, the shape of the SNARC effect seems to be similar in earlier and later phases within each task.

Exploratory 6: SNARC Comparison Between Tasks. A two-sided Bayesian paired *t*-test revealed strong evidence against a difference in strength of the continuous MC-SNARC (slope = -3.97 ms) and PJ-SNARC (slope = -3.57 ms) with BF_{10} = 0.06.

Exploratory 7: SNARC Correlation Between Tasks. Pearson's correlation between the MC-SNARC and the PJ-SNARC (using the better predictor for each task across participants as according to Hypotheses 1 and 2, i.e., categorical in MC and continuous in PJ) was $r = 0.17$. Note that the maximum attainable correlation equals the square root of the product of both measures' reliabilities (r_{MC} = 0.94 and r_{PJ} = 0.44, see Exploratory 4), yielding r_{max} = 0.64. A two-sided Bayesian Pearson correlation test (using the default r = 0.333 as a prior) provided strong evidence for this positive correlation with BF_{10} = 423.87.

Discussion

The present study aimed at replicating the SNARC effect, which is a characteristic effect in the field of numerical cognition, in large-scale online studies and at comparing the two tasks most commonly used to assess the SNARC effect. Although the SNARC effect is highly robust and generalizable across many settings, at least when running the standard continuous linear regression and averaging across all participants, it also depends on the experimental situation and task (Cipora, Patro, et al., 2018). The most important goal was to compare the shape of the SNARC effect between MC, where participants decide whether numbers are smaller or larger than 5, and PJ, where participants make judgments about them being odd or even. Moreover, a within-subjects design with high statistical power allowed for testing multiple other effects and task differences, as well as the correlation of SNARC effects observed in these

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

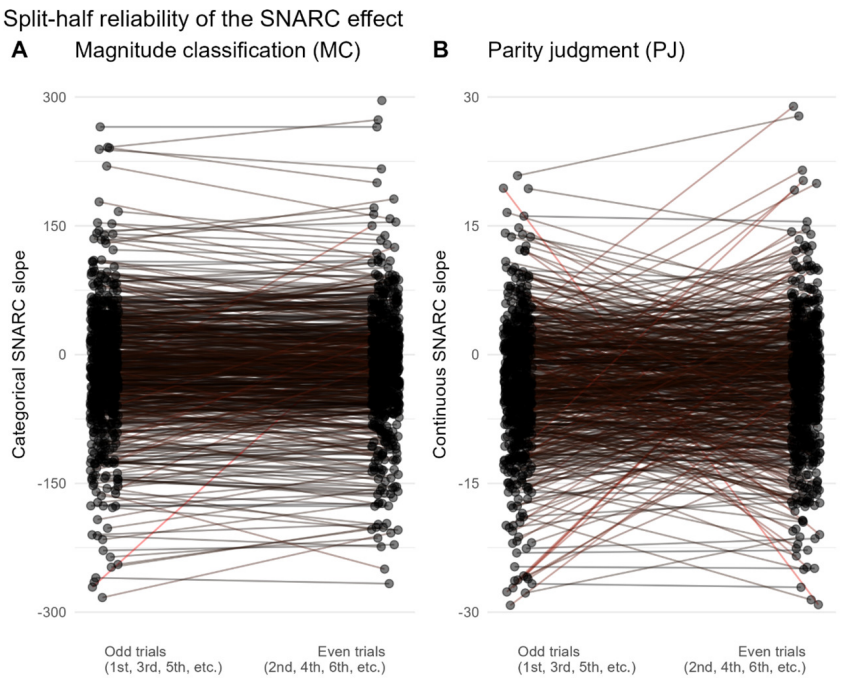


Figure 7. Plots illustrating the split-half reliability of the categorical MC-SNARC (Panel A) and the continuous PJ-SNARC (Panel B) (Exploratory 4)

Note. These plots illustrate the SNARC slopes for each participant as black points when calculated separately for odd and even trial numbers in the presentation order (i.e., 1st, 3rd, 5th, etc. trial vs. 2nd, 4th, 6th, etc. trial), reflecting the split-half reliability as by the odd-even method. The better fitting model according to results regarding Hypotheses 1 and 2 was used in each task (categorical MC-SNARC in Panel A and continuous PJ-SNARC in Panel B). Flatter lines connecting each participant's SNARC slopes between trial types (more black) speak for higher split-half reliability, whereas more diagonal ones (more red) speak for lower split-half reliability (with the color gradient being adjusted depending on relative steepness of the lines considering the range from -300 to 300 in MC and -30 to 30 in PJ).

two tasks. In the following, findings regarding the replication checks are summarized and discussed first, results regarding our three main hypotheses next, and observations from exploratory tests thereafter. Lastly, we give a summary and draw a conclusion with practical and theoretical implications.

Findings Obtained from the Replication Checks

Most observations from previous studies could be replicated: We found a continuous SNARC effect in both MC and PJ (Replication Check 1; for a comparison, see Exploratory 6), responses were faster in MC than in PJ (Replication Check 3), and there was a Numerical Distance Effect in MC (Replication Check 4). However, the data revealed strong evidence against the MARC effect in both tasks. Typically, the MC-MARC is not observed (Cipora, 2014; Deng et al., 2018), because the processing of parity and its relation to space is not as robust as that of magnitude (Roth, Caffier, Reips, Cipora, et al., 2025). In other words, the SNARC effect has been shown to be more automatic than the MARC effect. It is therefore expected not to find a MARC effect in MC, where the instructions do not address parity at all. However, it is surprising that we could not detect the PJ-MARC. Notably, the MARC effect was originally found in tasks with more emphasis on language, so it is more robust when using number words instead of Arabic numbers (Nuerk et al., 2004). Although in English, the term "odd" is not the literal negation of the term "even", so that linguistic markedness of parity might be weak, no difference in the PJ-MARC was found in an earlier comparison between languages with and without this negation (Cipora, Soltanlou, et al., 2019).

Findings Obtained from the Main Analysis

The main aim of the current study was to investigate the shape of the SNARC effect in the two most common tasks. As suggested in the existing literature (e.g., Wood et al., 2008), categorical magnitude in terms of numbers 1 to 4 being classified as "small" and numbers 6 to 9 being classified as "large" was a better predictor for the MC-SNARC than continuous magnitude from 1 to 9 (Hypothesis 1). However, this seems not to be the whole story, because datapoints revealed a negative slope instead of a horizontal alignment within small and large numbers when tested separately (Exploratory 17). This points towards an overall categorical MC-SNARC with a continuous shape within both small and large numbers (i.e., the truly observed pattern visible in Panel B in Figure 4 looks most like the illustration in Panel C in Figure 1). Note that this was previously not reported for the MC-SNARC in the literature. For instance, when summarizing studies included in an early meta-analysis, the MC-SNARC looked purely categorical with horizontally arranged datapoints within small and large numbers (see Figure 1 in Wood et al., 2008). One of our own studies (Bulut et al., 2025), however, revealed a categorical step in dRT between small and large numbers with a seemingly nega-

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

tive slope in small but not large numbers in all three samples. It is thus surprising that the present study speaks for an overall categorical shape with continuous aspects in both small and large numbers. Regarding the PJ-SNARC, as expected from the existing literature, the continuous predictor revealed a better fit than the categorical one (Hypothesis 2). The findings are in line with a fine-graded mental representation in PJ, where magnitude is automatically processed as a byproduct, as outlined in the Introduction, and with a more gross classification in MC, where magnitude is intentionally processed and classified, although the gross classification seems not to fully overrule the more fine-graded representation. Moreover, the findings partly support Gevers, Verguts, et al.'s (2006) theory that the combination of a Numerical Distance Effect (strong evidence in Exploratory 4) and a larger SNARC effect for slower responses (inconclusive evidence in Exploratory 18) lead to a more categorical MC-SNARC. Note that the fit of the categorical vs. continuous models did not differ much within each task in absolute terms ($R^2_{continous}$ = 0.421 vs. $R^2_{categorical}$ = 0.470 in MC, and $R^2_{continous}$ = 0.194 vs. $R^2_{categorical}$ = 0.164 in PJ), which might partly be due to the high collinearity (r = .913) between the categorical and continuous predictors. Interestingly, the proportion of explained variance R^2 is much larger in MC than in PJ in either model (see fit of the regression lines in Figure 4 vs. Figure 5). In the literature, a large range of reported R^2 can be found (ranging from 0.14 to 0.93 for the continuous and from 0.57 to 0.99 for the categorical MC-SNARC in adult participants, see Table A1 in Appendix A).

A Numerical Size Effect was also detected in both tasks, as expected (Hypothesis 3a). However, while there was strong evidence for it in PJ with the minimal sample size of 500 participants already, moderate evidence for it in MC was only found after increasing the sample size in four steps of 50 participants each, so that this optional recruitment stopping criterion was only met at a sample size of 700. Unexpectedly, the Numerical Size Effect was stronger in PJ than in MC (Hypothesis 3b). We had hypothesized that the effect would be stronger in MC, where processing magnitude was task-relevant, than in PJ, where it was task-irrelevant. A potential explanation for this result is that responses are generally faster in MC than in PJ, and the Numerical Size Effect has been shown to be stronger in slower responses (Hohol et al., 2020; see Supplementary Material, Table S3). We therefore checked separately in each task whether a correlation between the Numerical Size Effect and overall response speed could be found; however, our data revealed strong evidence against such a correlation (Exploratory 13, Appendix C), ruling out this explanation. A different potential explanation is that all number magnitudes should be paid attention to equally strongly when performing MC, attenuating the Numerical Size Effect, whereas number magnitude processing is not intentional but rather automatic in PJ. If a stronger neural overlap exists for large numbers, making small numbers easier to discriminate (Dehaene, 2001), this asymmetry may show via a stronger Numerical Size Effect when number magnitude processing is just a byproduct of the task, like in PJ.

Findings Obtained from the Exploratory Analysis

First, we found no task-order effects (Exploratory 1), at least when averaging across participants, which is a relevant finding for future studies in which both MC and PJ in within-subjects designs will be used. Evidence regarding compatibility-order effects on the PJ were inconclusive (Exploratory 2), being in line with null effects reported in several studies (Bulut et al., 2025; Cipora, van Dijck, et al., 2019; Moorkens et al., 2025; Roth et al., 2024). However, we detected compatibility-order effects showing that the MC-SNARC was stronger when participants completed the incompatible block first and the compatible block second, than when participants did the task in the reversed order (in line with findings by Bulut et al., 2025, but opposed to findings in children by Van Galen & Reitsma, 2008). This is a typical pattern, because when the first block is incompatible, participants are particularly slow for two reasons (see Roth et al., 2024): They have to familiarize themselves with the task (Exploratory 8, Appendix C), and they additionally need to overcome their natural response bias due to the SNARC effect. When the first block is compatible, however, RT are more similar between blocks, because participants familiarize themselves with the task in the first block, whereas they need to overcome their natural response bias due to the SNARC effect in the second block. Because there were only few dropouts, we did not check for motivational confounding (Reips, 2002; Reips et al., 2025), which may have resulted from the compatibility-order asymmetry. For compatibility effects on general RT, please refer to Exploratory 11 and 12 (Appendix C).

As discussed earlier, the SNARC effect is not necessarily purely categorical or continuous (see Panel C in Figure 1), especially in MC. Evidence favored the continuous predictor for MC in models including both categorical and continuous predictors (Exploratories 3 and 5; note limited interpretability due to high collinearity between predictors and due to the existence of only eight data points). In contrast, the categorical fit was slightly higher than the continuous fit for MC when comparing two separate models (Hypothesis 1), suggesting that the shape of the SNARC effect is sensitive to the analysis used. While researchers may use a simple categorical or continuous model to quantify the strength of the MC-SNARC or PJ-SNARC, respectively (e.g., for between-group comparisons or testing the influence of experimental manipulations on the SNARC effect), we advise them to analyze the exact shape of the MC-SNARC with the following steps: First, they should compute a categorical model (Hypothesis 1), and second, if visual inspection of dRT suggests a continuous shape within each half of the stimulus set, they should additionally test continuous slopes separately for small and large numbers (Exploratory 17).

Regarding the categorization into “small” and “large” numbers, we compared the fit of five regression models using different boundaries for each task (Exploratory 4). We suspected that number 5 would be the boundary in most participants in MC, where it is explicitly defined as such according to the task instructions, while there might not

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

be any clearcut boundary in PJ. We found that number 5 was indeed the most likely underlying boundary for about half of the sample in MC. Adjacent boundaries (3.5 and 6.5) were descriptively still more prominent than most extreme boundaries (2.5 and 7.5), leading to a unimodal distribution around 5 in MC (see Table 8). In contrast, the distribution of boundaries seemed rather random in the analysis for PJ, so that descriptively, number 5 even happened to be the least favored boundary within the sample. However, not only does the most likely underlying boundary vary considerably between individuals, especially in PJ, but also it seems to be rather unstable within individuals (see Figure 6): The split-half reliability of the boundaries could not be distinguished from 0 in either of the tasks. Note that the stability was rather high for those participants revealing the boundary 5 in MC (see Figure 6). Expectedly, when comparing the continuous model with the categorical model using the favored boundary for each participant, the categorical model still fits the data better than the continuous one in MC (as in Hypothesis 1). More interestingly, this was also the case for PJ, suggesting that the underlying shape of the spatial number representation might be categorical in that task as well, and just vary a lot more between participants. One could therefore think that the better fit of the continuous model found in the main analysis (Hypothesis 2) might only result from large interindividual differences in categorical boundaries and averaging across various favored models. However, taking into account the null split-half reliability discussed above and often dramatic changes in categorical boundaries in PJ between the halves suggests that the underlying model is in fact continuous.

The split-half reliability of the MC-SNARC was high (r = 0.94), whereas the split-half reliability of the PJ-SNARC was lower (r = 0.44), and also at the lower end of the range previously reported in the literature (Exploratory 4, Figure 7). These differences between tasks might be caused by the MC-SNARC resulting from an aggregation of only two possible values for the magnitude predictor, whereas the PJ-SNARC results from eight datapoints.

When quantified with a continuous predictor (although a categorical one fits better in MC) the MC-SNARC and the PJ-SNARC were similarly strong (Exploratory 6) and weakly correlated with one another (Exploratory 7). A stronger MC-SNARC than PJ-SNARC would have been expected in light of stronger magnitude processing in MC than in PJ: In terms of the dual-route model by Gevers, Ratinckx, et al. (2006), both the slow conditional and the fast unconditional route are activated in MC, but only the fast unconditional route is activated in PJ. In other words, one of both prerequisites for the SNARC (as defined by Cipora, He, et al., 2020) is fulfilled in MC anyways because of the task instructions, namely magnitude processing, so that only the other one needs to happen automatically, namely the directional association. However, the opposite prediction (i.e., a stronger PJ-SNARC than MC-SNARC) could be derived from longer RT in PJ than in MC and a typically stronger SNARC in slower responses. Although evidence for the latter remained inconclusive despite the large amount of data, the direction of the relationship was as expected (Exploratory 18). As outlined in the Introduction, both reasonings are plausible, and the observed lack of a difference between the MC-SNARC and the PJ-SNARC might be due to both mechanisms cancelling each other out. The positive correlation between the MC-SNARC and the PJ-SNARC is interesting in light of the (yet underpowered) literature, where, to our knowledge, a correlation had only been found to be significant by Cheung et al. (2015) and by Cipora (2014; but only in a unimanual setup), while null results were obtained in all other studies (Bulut et al., 2025; Cipora, 2014, and Cipora et al., 2016, for PJ, and Hohol et al., 2020, for MC; Didino et al., 2019; Fattorini et al., 2015, Study 1; Georges et al., 2017; Van Dijck & Doricchi, 2019). For an overview, please refer to the rightmost column in Table A1 (Appendix A). Arguments against a correlation might be low intraindividual stability, low reliability, or low internal consistency of the SNARC effect in both tasks. However, the results of the present study might be more trustworthy thanks to the large sample size, and previous studies might just have failed to detect an existing but small correlation. The observed correlation leads to the conclusion that the mental representation of numbers underlying in MC and PJ overlaps at least partly.

In addition to what was proposed in the Stage-1 Registered Report, the large amount of data permitted us to investigate several other aspects of MC and PJ (see Appendix C). When habituating to such tasks, participants usually get faster (i.e., training effects). This was reflected by faster responses in the second than in the first block for the case of MC (Exploratory 8, Appendix C), and by faster responses in the second than first half of each block for the case of PJ (Exploratory 9, Appendix C).

A typical Odd effect (i.e., faster responses to even than to odd numbers) was only observed in PJ (Exploratory 10, Appendix C). A reversed Odd effect was observed in MC, which seems striking at first sight. However, this is fully in line with the large Numerical Distance Effect observed in MC (Replication Check 4): The numbers closest to the reference number 5 (i.e., 4 and 6) are responded to more slowly than the numbers with the greatest numerical distance to the reference number 5 (i.e., 1 and 9); and crucially, the first happen to be even, whereas the latter happen to be odd. Thus, it seems that the Numerical Distance Effect is overruling the Odd effect, which might not be as strong in MC as in PJ anyways (see Figure 3).

A typical effect on overall RT was observed regarding the SNARC-compatibility in each task, such that responses were faster for SNARC-compatible than for SNARC-incompatible trials (by 10 ms in MC and by 8 ms in PJ, on average; Exploratory 11, Appendix C). No such MARC-compatibility effect on overall RT could be found, with RT being similar for MARC-compatible and MARC-incompatible trials in each task (Exploratory 12, Appendix C). This null result (in Bayesian sense) is in line with the lack of a MARC effect in both tasks (Replication Check 2), and also with the lack of a compatibility-order effect on the PJ-MARC, where MARC-compatible and MARC-incompatible trials were presented in two separate blocks (Exploratory 2). Note that the null

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

MARC effect did not differ between left- and right-handers (Exploratory 16, Appendix C).

Lastly, Moretti et al. (2025) recently suggested to focus on trials that elicit the highest response conflict when measuring cognitive effects. We therefore re-analyzed the PJ-SNARC and re-checked its split-half reliability while only considering post-SNARC-compatible trials (i.e., following trials with SNARC-congruent stimulus presentations: 1, 3, 6, 8 in blocks with odd-left/even-right instructions, or 2, 4, 7, 9 in blocks with even-left/odd-right instructions). The effect size of the continuous PJ-SNARC was much larger when excluding post-SNARC-incompatible trials (Exploratory 14, Appendix C) than when keeping them (Hypothesis 2); however, its split-half reliability remained similar (Exploratory 15, Appendix C).

Summary and Conclusion

In this large-scale online experiment, the SNARC effect was replicated in the two most commonly used tasks (namely, MC and PJ). Analyzing data from over 600 participants, a rather categorical shape of the MC-SNARC with additional continuous tendencies and a rather continuous shape of the PJ-SNARC were found. Assessing these two most widely used SNARC tasks in a within-subjects design also permitted to detect a small but positive correlation between the MC-SNARC and the PJ-SNARC. Moreover, we thoroughly investigated numerous other numerical-cognition and general-RT phenomena, laying a comprehensive groundwork for further investigations using MC and PJ.

The most important methodological implication of this work for scientific practice, given that statistical models should align with scientific ones (Westermann & Hager, 1983), is that researchers should use a categorical statistical model to quantify the SNARC effect in MC. They might even want to consider computing different categorical models with individual boundaries between "small" and "large" numbers, and additionally investigate continuous aspects. At the same time, researchers can continue using the traditional continuous model in PJ. The most important theoretical implication of this work is that, instead of speaking about *the* SNARC effect, researchers should rather acknowledge that the shape and other properties of the SNARC effects depend on the used task, and the underlying roots for different types of SNARC effects may be different as well.

Author Contributions

Data curation: Lilly Roth, Krzysztof Cipora, Annika Tave Overlander, Hans-Christoph Nuerk, Ulf-Dietrich Reips. Formal Analysis: Lilly Roth, Krzysztof Cipora, Annika Tave Overlander, Hans-Christoph Nuerk, Ulf-Dietrich Reips. Investigation: Lilly Roth, Krzysztof Cipora, Annika Tave Overlander, Hans-Christoph Nuerk, Ulf-Dietrich Reips. Methodology: Lilly Roth, Krzysztof Cipora, Hans-Christoph Nuerk, Ulf-Dietrich Reips. Validation: Lilly Roth, Krzysztof Cipora, Annika Tave Overlander, Hans-Christoph Nuerk, Ulf-Dietrich Reips. Visualization: Lilly Roth. Writing – original draft: Lilly Roth. Conceptualization: Krzysztof Cipora, Hans-Christoph Nuerk, Ulf-Dietrich Reips. Funding acquisition: Krzysztof Cipora, Hans-Christoph Nuerk, Ulf-Dietrich Reips. Supervision: Krzysztof Cipora, Hans-Christoph Nuerk, Ulf-Dietrich Reips. Writing – review & editing: Krzysztof Cipora, Annika Tave Overlander, Hans-Christoph Nuerk, Ulf-Dietrich Reips. Software: Annika Tave Overlander, Ulf-Dietrich Reips. Project administration: Hans-Christoph Nuerk, Ulf-Dietrich Reips. Resources: Hans-Christoph Nuerk, Ulf-Dietrich Reips.

Competing Interests

The authors declare no conflicts of interest with the content of this article.

Funding

This research was supported by the DFG project "Replicability of Fundamental Results on Spatial-Numerical Associations in Highly Powered Online Experiments (e-SNARC)" (NU 265/8-1 and RE 2655/3-1) granted to Hans-Christoph Nuerk and Ulf-Dietrich Reips, supporting Lilly Roth and Annika Tave Overlander, with the assistance of Krzysztof Cipora as a cooperation partner. Hans-Christoph Nuerk's work on Spatial-Numerical Associations is additionally supported by the DFG-projects 265-5/1 and 265-5/2: On the interplay of modal and amodal encodings underlying space-metric associations (SMAs). Krzysztof Cipora is supported by the UKRI Economic and Social Research Council (grant number ES/W002914/1).

Acknowledgements

The authors would like to thank Sebastian Sandbrink for proofreading the Stage-1 Registered Report.

Data and Code Availability

All R scripts for data preprocessing and analysis as well as all anonymized datasets were made publicly available via the Open Science Framework (https://osf.io/g48s2/) and Zenodo (https://doi.org/10.5281/zenodo.22140752), along with all previous manuscript versions.

Appendices files can be downloaded from the on-line version

Editors: Don van Ravenzwaaij (Editor-in-Chief)

Submitted: July 23, 2026 PDT. Accepted: August 19, 2026 PDT.
Published: October 22, 2026 PDT.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at http://creativecommons.org/licenses/by/4.0 and legal code at http://creativecommons.org/licenses/by/4.0/legalcode for more information.

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

References

Bachot, J., Gevers, W., Fias, W., & Roeyers, H. (2005). Number sense in children with visuospatial disabilities: Orientation of the mental number line. *Psychology Science*, *47*(1), 172–183.

Bae, G. Y., Choi, J. M., Cho, Y. S., & Proctor, R. W. (2009). Transfer of magnitude and spatial mappings to the SNARC effect for parity judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1506–1521. https://doi.org/10.1037/a0017257

Banks, W. P., Fujii, M., & Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 435–447. https://doi.org/10.1037/0096-1523.2.3.435

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Hocking, T., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bonsch, M., Parsonage, H., Ritchie, S., Ren, K., Tan, X., Saporta, R., Seiskari, O., Dong, X., Lang, M., … Czekanski, M. (2024). *data.table: Extension of "data.frame"* [R package, version 1.17.0]. https://cran.r-project.org/web/packages/data.table/index.html

Basso Moro, S., Dell'Acqua, R., & Cutini, S. (2018). The SNARC effect is not a unitary phenomenon. *Psychonomic Bulletin & Review*, *25*(2), 688–695. https://doi.org/10.3758/s13423-017-1408-3

Brunson, J. C., & Read, Q. D. (2023). *ggalluvial: Alluvial plots in "ggplot2"* [R package, version 0.12.5]. https://CRAN.R-project.org/package=ggalluvial

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General*, *124*(4), 434–452. https://doi.org/10.1037/0096-3445.124.4.434

Buerkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bull, R., Marschark, M., & Blatto-Vallee, G. (2005). SNARC hunting: Examining number representation in deaf students. *Learning and Individual Differences*, *15*(3), 223–236. https://doi.org/10.1016/j.lindif.2005.01.004

Bulut, M., Roth, L., Bahreini, N., Cipora, K., Reips, U.-D., & Nuerk, H.-C. (2025). One direction? Cultural aspects of the mental number line beyond reading direction. *Psychological Research*, *89*, Article 37. https://doi.org/10.1007/s00426-024-02038-4

Casasanto, D., & Pitt, B. (2019). The faulty magnitude detector: Why SNARC-like tasks cannot support a generalized magnitude system. *Cognitive Science*, *43*(10), Article e12794. https://doi.org/10.1111/cogs.12794

Cheung, C.-N., Ayzenberg, V., Diamond, R. F. L., Yousif, S. R., & Lourenco, S. F. (2015). Probing the mental number line: A between-task analysis of spatial-numerical associations. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 357–362.

Cipora, K. (2014). *Between-task consistency of the SNARC effect* [Poster presentation]. 33rd European Workshop on Cognitive Neuropsychology, Bressanone, Italy. https://osf.io/nx6vq/

Cipora, K., & Göbel, S. M. (2013). Number-space associations: Just how reliable is the SNARC effect [Poster presentation]. In *32nd European Workshop on Cognitive Neuropsychology*, Bressanone, Italy. https://osf.io/nx6vq

Cipora, K., Haman, M., Domahs, F., & Nuerk, H.-C. (2020). Editorial: On the development of space-number relations: Linguistic and cognitive determinants, influences, and associations. *Frontiers in Psychology*, *11*, Article 182. https://doi.org/10.3389/fpsyg.2020.00182

Cipora, K., He, Y., & Nuerk, H.-C. (2020). The spatial–numerical association of response codes effect and math skills: Why related? *Annals of the New York Academy of Sciences*, *1477*(1), 5–19. https://doi.org/10.1111/nyas.14355

Cipora, K., Patro, K., & Nuerk, H.-C. (2018). Situated influences on spatial–numerical associations. In T. L. Hubbard (Ed.), *Spatial biases in perception and cognition* (pp. 41–59). Cambridge University Press. https://doi.org/10.1017/9781316651247.004

Cipora, K., Schroeder, P. A., Soltanlou, M., & Nuerk, H.-C. (2018). More space, better mathematics: Is space a powerful tool or a cornerstone for understanding arithmetic? In K. Mix & M. Batista (Eds.), *Visualizing mathematics: The role of spatial reasoning in mathematical thought* (pp. 77–116). Springer. https://doi.org/10.1007/978-3-319-98767-5_4

Cipora, K., Soltanlou, M., Reips, U.-D., & Nuerk, H.-C. (2019). The SNARC and MARC effects measured online: Large-scale assessment methods in flexible cognitive effects. *Behavior Research Methods*, *51*(4), 1676–1692. https://doi.org/10.3758/s13428-019-01213-5

Cipora, K., van Dijck, J.-P., Georges, C., Masson, N., Goebel, S. M., Willmes, K., Pesenti, M., Schiltz, C., & Nuerk, H.-C. (2019). *A minority pulls the sample mean: On the individual prevalence of robust group-level cognitive phenomena – the instance of the SNARC effect*. Open Science Framework. https://doi.org/10.31234/osf.io/bwyr3

Cipora, K., van Dijck, J.-P., Georges, C., Masson, N., Schiltz, C., Pesenti, M., Goebel, S. M., Willmes, K., & Nuerk, H.-C. (2020). *Parity and space: On the prevalence and individual differences in the MARC effect* [Poster presentation]. 38th European Workshop on Cognitive Neuropsychology, Bressanone, Italy. https://osf.io/4qhpw

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Cipora, K., & Wood, G. (2017). Finding the SNARC instead of hunting it: A 20∗ 20 Monte Carlo investigation. *Frontiers in Psychology*, *8*, Article 1194. https://doi.org/10.3389/fpsyg.2017.01194

Correll, J., Mellinger, C., McClelland, G. H., & Judd, C. M. (2020). Avoid Cohen's "small", "medium", and "large" for power analysis. *Trends in Cognitive Sciences*, *24*(3), 200–207. https://doi.org/10.1016/j.tics.2019.12.009

Dehaene, S. (2001). Subtracting pigeons: Logarithmic or linear? *Psychological Science*, *12*(3), 244–246. https://doi.org/10.1111/1467-9280.00343

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Deng, Z., Chen, Y., Zhang, M., Li, Y., & Zhu, X. (2018). The association of number and space under different tasks: Insight from a process perspective. *Frontiers in Psychology*, *9*, Article 957. https://doi.org/10.3389/fpsyg.2018.00957

Deng, Z., Chen, Y., Zhu, X., & Li, Y. (2017). The effect of working memory load on the SNARC effect: Maybe tasks have a word to say. *Memory & Cognition*, *45*(3), 428–441. https://doi.org/10.3758/s13421-016-0676-x

Didino, D., Breil, C., & Knops, A. (2019). The influence of semantic processing and response latency on the SNARC effect. *Acta Psychologica*, *196*, 75–86. https://doi.org/10.1016/j.actpsy.2019.04.008

Dienes, Z. (2021). How to use and report Bayesian hypothesis tests. *Psychology of Consciousness: Theory, Research, and Practice*, *8*(1), 9–26. https://doi.org/10.1037/cns0000258

Fattorini, E., Pinto, M., Rotondaro, F., & Doricchi, F. (2015). Perceiving numbers does not cause automatic shifts of spatial attention. *Cortex*, *73*, 298–316. https://doi.org/10.1016/j.cortex.2015.09.007

Fias, W., Brysbaert, M., Geypens, F., & D'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition*, *2*(1), 95–110. https://doi.org/10.1080/135467996387552

Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition: From single digits to arithmetic. *Quarterly Journal of Experimental Psychology*, *67*(8), 1461–1483. https://doi.org/10.1080/17470218.2014.927515

Fitousi, D., Shaki, S., & Algom, D. (2009). The role of parity, physical size, and magnitude in numerical cognition: The SNARC effect revisited. *Attention, Perception & Psychophysics*, *71*(1), 143–155. https://doi.org/10.3758/APP.71.1.143

Garaizar, P., & Reips, U.-D. (2019). Best practices: Two web browser-based methods for stimulus presentation in behavioral experiments with high resolution timing requirements. *Behavior Research Methods*, *51*, 1441–1453. https://doi.org/10.3758/s13428-018-1126-4

Georges, C., Hoffmann, D., & Schiltz, C. (2013). *The SNARC effect and its relationship to spatial abilities in women* [Poster presentation]. 18th conference of the European Society for Cognitive Psychology, Budapest, Hungary. http://hdl.handle.net/10993/13012

Georges, C., Hoffmann, D., & Schiltz, C. (2017). How and why do number-space associations co-vary in implicit and explicit magnitude processing tasks? *Journal of Numerical Cognition*, *3*(2), 182–211. https://doi.org/10.5964/jnc.v3i2.46

Gevers, W., Ratinckx, E., de Baene, W., & Fias, W. (2006). Further evidence that the SNARC effect is processed along a dual-route architecture: Evidence from the lateralized readiness potential. *Experimental Psychology*, *53*(1), 58–68. https://doi.org/10.1027/1618-3169.53.1.58

Gevers, W., Santens, S., Dhooge, E., Chen, Q., van den Bossche, L., Fias, W., & Verguts, T. (2010). Verbal-spatial and visuospatial coding of number-space interactions. *Journal of Experimental Psychology: General*, *139*(1), 180–190. https://doi.org/10.1037/a0017688

Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 32–44. https://doi.org/10.1037/0096-1523.32.1.32

Gökaydin, D., Brugger, P., & Loetscher, T. (2018). Sequential effects in SNARC. *Scientific Reports*, *8*(1), Article 10996. https://doi.org/10.1038/s41598-018-29337-2

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480–506. https://doi.org/10.1037/0096-3445.121.4.480

Han, M., Mao, X., Cai, M., Jia, X., & Guo, C. (2017). The effect of positive and negative signs on the SNARC effect in the magnitude judgment task. *Acta Psychologica Sinica*, *49*(8), 995–1008. https://doi.org/10.3724/SP.J.1041.2017.00995

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Herrera, A., Macizo, P., & Semenza, C. (2008). The role of working memory in the association between number magnitude and space. *Acta Psychologica*, *128*(2), 225–237. https://doi.org/10.1016/j.actpsy.2008.01.002

Hines, T. M. (1990). An odd effect: Lengthened reaction times for judgments about odd digits. *Memory & Cognition*, *18*(1), 40–46. https://doi.org/10.3758/BF03202644

Hoffmann, D., Hornung, C., Martin, R., & Schiltz, C. (2013). Developing number-space associations: SNARC effects using a color discrimination task in 5-year-olds. *Journal of Experimental Child Psychology*, *116*(4), 775–791. https://doi.org/10.1016/j.jecp.2013.07.013

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Hohol, M., Willmes, K., Nęcka, E., Brożek, B., Nuerk, H.-C., & Cipora, K. (2020). Professional mathematicians do not differ from others in the symbolic numerical distance and size effects. *Scientific Reports*, *10*(1), Article 11531. https://doi.org/10.1038/s41598-020-68202-z

Hohol, M., Wołoszyn, K., & Cipora, K. (2022). No fingers, no SNARC? Neither the finger counting starting hand, nor its stability robustly affect the SNARC effect. *Acta Psychologica*, *230*, Article 103765. https://doi.org/10.1016/j.actpsy.2022.103765

Hubbard, E. M., Ranzini, M., Piazza, M., & Dehaene, S. (2009). What information is critical to elicit interference in number-form synaesthesia? *Cortex*, *45*(10), 1200–1216. https://doi.org/10.1016/j.cortex.2009.06.011

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, *32*(4), 662–673. https://doi.org/10.3758/BF03195857

Kassambara, A. (2025). *ggpubr: "ggplot2" based publication ready plots* [R package, version 0.6.1]. https://CRAN.R-project.org/package=ggpubr

Kelter, R. (2021). Analysis of type I and II error rates of Bayesian and frequentist parametric and nonparametric two-sample hypothesis tests under preliminary assessment of normality. *Computational Statistics*, *36*(2), 1263–1288. https://doi.org/10.1007/s00180-020-01034-7

Kiesel, A., Wendt, M., & Peters, A. (2007). Task switching: On the origin of response congruency effects. *Psychological Research*, *71*(2), 117–125. https://doi.org/10.1007/s00426-005-0004-8

Koch, N. N., Huber, J. F., Lohmann, J., Cipora, K., Butz, M. V., & Nuerk, H.-C. (2023). Mental number representations are spatially mapped both by their magnitudes and ordinal positions. *Collabra: Psychology*, *9*(1), Article 67908. https://doi.org/10.1525/collabra.67908

Kopiske, K. K., Löwenkamp, C., Eloka, O., Schiller, F., Kao, C.-S., Wu, C., Gao, X., & Franz, V. H. (2016). The SNARC effect in Chinese numerals: Do visual properties of characters and hand signs influence number processing? *PLOS ONE*, *11*(9), Article e0163897. https://doi.org/10.1371/journal.pone.0163897

Lohmann, J., Schroeder, P. A., Nuerk, H.-C., Plewnia, C., & Butz, M. V. (2018). How deep is your SNARC? Interactions between numerical magnitude, response hands, and reachability in peripersonal space. *Frontiers in Psychology*, *9*, Article 622. https://doi.org/10.3389/fpsyg.2018.00622

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 149–157. https://doi.org/10.1037/0278-7393.16.1.149

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Moorkens, J., Dijck, J.-P. van, & Fias, W. (2025). The parity judgment SNARC effect: The role of response mapping order and the nature of the instruction. *Journal of Numerical Cognition*, *11*, Article e15051. https://doi.org/10.5964/jnc.15051

Moretti, L., Koch, I., Hornjak, R., & von Bastian, C. C. (2025). Quality over quantity: Focusing on high-conflict trials to improve the reliability and validity of attentional control measures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001466

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2024). *BayesFactor: Computation of Bayes factors for common designs*. R package, version 0.9.12-4.7. https://CRAN.R-project.org/package=BayesFactor

Mourad, A., & Leth-Steensen, C. (2017). Spatial reference frames and SNARC. *Journal of Cognitive Psychology*, *29*(2), 113–128. https://doi.org/10.1080/20445911.2016.1249483

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520. https://doi.org/10.1038/2151519a0

Nathan, M. B., Shaki, S., Salti, M., & Algom, D. (2009). Numbers and space: Associations and dissociations. *Psychonomic Bulletin & Review*, *16*(3), 578–582. https://doi.org/10.3758/PBR.16.3.578

Nuerk, H.-C., Bauer, F., Krummenacher, J., Heller, D., & Willmes, K. (2005). The power of the mental number line: How the magnitude of unattended numbers affects performance in an Eriksen task. *Psychology Science*, *47*(1), 34–50. https://psycnet.apa.org/record/2005-11470-005

Nuerk, H.-C., Iversen, W., & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Quarterly Journal of Experimental Psychology*, *57*(5), 835–863. https://doi.org/10.1080/02724980343000512

Nuerk, H.-C., & Schroeder, P. A. (2024). *A Psychological Markedness Account of human cognition: The example of numerical cognition* [Manuscript in preparation]. Department of Psychology, University of Tuebingen, Germany.

Nuerk, H.-C., Wood, G., & Willmes, K. (2005). The universal SNARC effect: The association between number magnitude and space is amodal. *Experimental Psychology*, *52*(3), 187–194. https://doi.org/10.1027/1618-3169.52.3.187

Patro, K., Nuerk, H. C., Cress, U., & Haman, M. (2014). How number-space relationships are assessed before formal schooling: A taxonomy proposal. *Frontiers in Psychology*, *5*, Article 419. https://doi.org/10.3389/fpsyg.2014.00419

Pfister, R., Schroeder, P. A., & Kunde, W. (2013). SNARC struggles: Instant control over spatial-numerical associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1953–1958. https://doi.org/10.1037/a0032991

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Pinto, M., Pellegrino, M., Lasaponara, S., Scozia, G., D'Onofrio, M., Raffa, G., Nigro, S., Arnaud, C. R., Tomaiuolo, F., & Doricchi, F. (2021). Number space is made by response space: Evidence from left spatial neglect. *Neuropsychologia*, *154*, Article 107773. https://doi.org/10.1016/j.neuropsychologia.2021.107773

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416

Prpic, V., Fumarola, A., de Tommaso, M., Luccio, R., Murgia, M., & Agostini, T. (2016). Separate mechanisms for magnitude and order processing in the spatial-numerical association of response codes (SNARC) effect: The strange case of musical note values. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1241–1251. https://doi.org/10.1037/xhp0000217

R Core Team. (2024). *R: A language and environment for statistical computing* (4.4.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Reips, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 89–117). Academic Press. https://doi.org/10.1016/B978-012099980-4/50005-8

Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, *49*(4), 243–256. https://doi.org/10.1026//1618-3169.49.4.243

Reips, U.-D. (2007). The methodology of Internet-based experiments. In A. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 373–390). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199561803.013.0024

Reips, U.-D. (2009). Internet experiments: Methods, guidelines, metadata. In B. E. Rogowitz & T. N. Pappas (Eds.), *SPIE Proceedings, Human Vision and Electronic Imaging XIV*, Article 724008. International Society for Optics / SPIE. https://doi.org/10.1117/12.823416

Reips, U.-D. (2021). Web-based research in psychology: A review. *Zeitschrift für Psychologie*, *229*(4), 198–213. https://doi.org/10.1027/2151-2604/a000475

Reips, U.-D., & Neuhaus, C. (2002). WEXTOR: A Web-based tool for generating and visualizing experimental designs and procedures. *Behavior Research Methods, Instruments, & Computers*, *34*(2), 234–240. https://doi.org/10.3758/bf03195449

Reips, U.-D., Overlander, A. T., & Bannert, M. (2025). Dropout analysis: A method for data from Internet-based research and dropR, an R-based web app and package to analyze and visualize dropout. *Behavior Research Methods*, *57*(1), Article 231. https://doi.org/10.3758/s13428-025-02730-2

Roettger, T. B., & Domahs, F. (2015). Grammatical number elicits SNARC and MARC effects as a function of task demands. *Quarterly Journal of Experimental Psychology (2006)*, *68*(6), 1231–1248. https://doi.org/10.1080/17470218.2014.979843

Roth, L., Caffier, J., Reips, U.-D., Cipora, K., Braun, L., & Nuerk, H.-C. (2025). True colors SNARC: Semantic number processing is highly automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001431

Roth, L., Caffier, J., Reips, U.-D., Nuerk, H.-C., Overlander, A. T., & Cipora, K. (2025). One and only SNARC? Spatial-Numerical Associations are not fully flexible and depend on both relative and absolute number magnitude. *Royal Society Open Science*, *12*(1), Article 241585. https://doi.org/10.1098/rsos.241585

Roth, L., Huber, J. F., Kronenthaler, S., van Dijck, J., Cipora, K., Butz, M. V., & Nuerk, H.-C. (2025). Looks like SNARC spirit: Coexistence of short- and long-term associations between letters and space. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218251324437

Roth, L., Jordan, V., Schwarz, S., Willmes, K., Nuerk, H.-C., van Dijck, J.-P., & Cipora, K. (2024). Don't SNARC me now! Intraindividual variability of cognitive phenomena – Insights from the Ironman paradigm. *Cognition*, *248*, Article 105781. https://doi.org/10.1016/j.cognition.2024.105781

Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods*, *13*(1), 19–30. https://doi.org/10.1037/1082-989X.13.1.19

Saeki, E., & Saito, S. (2009). Verbal representation in task order control: An examination with transition and task cues in random task switching. *Memory & Cognition*, *37*(7), 1040–1050. https://doi.org/10.3758/MC.37.7.1040

Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition*, *108*(1), 263–270. https://doi.org/10.1016/j.cognition.2008.01.002

Schiller, F., Eloka, O., & Franz, V. H. (2016). Using key distance to clarify a theory on the SNARC. *Perception*, *45*(1–2), 196–221. https://doi.org/10.1177/0301006615616754

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schroeder, P. A., Nuerk, H.-C., & Plewnia, C. (2017). Switching between multiple codes of SNARC-like associations: Two conceptual replication attempts with anodal tDCS in sham-controlled cross-over design. *Frontiers in Neuroscience*, *11*, Article 654. https://doi.org/10.3389/fnins.2017.00654

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Perception & Psychophysics*, *66*(4), 651–664. https://doi.org/10.3758/bf03194909

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026

Schwarz, W., & Müller, D. (2006). Spatial associations in number-related tasks: A comparison of manual and pedal responses. *Experimental Psychology*, *53*(1), 4–15. https://doi.org/10.1027/1618-3169.53.1.4

Shaki, S., & Gevers, W. (2011). Cultural characteristics dissociate magnitude and ordinal information processing. *Journal of Cross-Cultural Psychology*, *42*(4), 639–650. https://doi.org/10.1177/0022022111406100

Tlauka, M. (2002). The processing of numbers in choice-reaction tasks. *Australian Journal of Psychology*, *54*(2), 94–98. https://doi.org/10.1080/00049530210001706553

Toomarian, E. Y., & Hubbard, E. M. (2018). On the genesis of Spatial-Numerical Associations: Evolutionary and cultural factors co-construct the Mental Number Line. *Neuroscience & Biobehavioral Reviews*, *90*, 184–199. https://doi.org/10.1016/j.neubiorev.2018.04.010

Tzelgov, J. (1997). Specifying the relations between automaticity and consciousness: A theoretical note. *Consciousness and Cognition*, *6*(2–3), 441–451. https://doi.org/10.1006/ccog.1997.0303

Tzelgov, J., Ganor-Stern, D., Kallai, A. Y., & Pinhas, M. (2015). Primitives and non-primitives of numerical representations. In R. C. Kadosh & A. Dowker (Eds.), *Oxford library of psychology: The Oxford handbook of numerical cognition* (pp. 45–66). Oxford University Press.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 166–179. https://doi.org/10.1037/0278-7393.18.1.166

Van Dijck, J.-P., Abrahamse, E. L., Acar, F., Ketels, B., & Fias, W. (2014). A working memory account of the interaction between numbers and spatial attention. *Quarterly Journal of Experimental Psychology*, *67*(8), 1500–1513. https://doi.org/10.1080/17470218.2014.903984

Van Dijck, J.-P., & Doricchi, F. (2019). Multiple left-to-right spatial representations of number magnitudes? Evidence from left spatial neglect. *Experimental Brain Research*, *237*(4), 1031–1043. https://doi.org/10.1007/s00221-019-05483-5

Van Dijck, J.-P., Gevers, W., & Fias, W. (2009). Numbers are associated with different types of spatial information depending on the task. *Cognition*, *113*(2), 248–253. https://doi.org/10.1016/j.cognition.2009.08.005

Van Dijck, J.-P., Gevers, W., Lafosse, C., & Fias, W. (2012). The heterogeneous nature of number-space interactions. *Frontiers in Human Neuroscience*, *5*, Article 182. https://doi.org/10.3389/fnhum.2011.00182

Van Galen, M. S., & Reitsma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology*, *101*(2), 99–113. https://doi.org/10.1016/j.jecp.2008.05.001

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Viarouge, A., Hubbard, E. M., & McCandliss, B. D. (2014). The cognitive mechanisms of the SNARC effect: An individual differences approach. *PLOS ONE*, *9*(4), Article e95756. https://doi.org/10.1371/journal.pone.0095756

Weis, T., Nuerk, H.-C., & Lachmann, T. (2018). Attention allows the SNARC effect to operate on multiple number lines. *Scientific Reports*, *8*(1), Article 13778. https://doi.org/10.1038/s41598-018-32174-y

Westermann, R., & Hager, W. (1983). On severe tests of trend hypotheses in psychology. *The Psychological Record*, *33*(2), 201–211. https://doi.org/10.1007/BF03394838

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., van den Brand, T., & Posit PBC. (2024). *ggplot2: Create elegant data visualisations using the grammar of graphics*. R Package, Version 3.5.2. https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A grammar of data manipulation*. R Package, Version 1.1.4. https://cran.r-project.org/web/packages/dplyr/index.html

Wickham, H., Pedersen, T. L., Seidel, D., & Posit Software. (2025). *scales: Scale Functions for Visualization*. R Package, Version 1.4.0. https://cran.r-project.org/web/packages/scales/index.html

Wickham, H., Vaughan, D., Girlich, M., & Ushey, K. (2023). *tidyr: Tidy messy data*. R Package, Version 1.3.1. https://cran.r-project.org/web/packages/tidyr/index.html

Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science*, *50*(4), 489–525.

Xiang, X., Yan, L., Fu, S., & Nan, W. (2022). Processing stage flexibility of the SNARC effect: Task relevance or magnitude relevance? *Frontiers in Psychology*, *13*, Article 1022999. https://doi.org/10.3389/fpsyg.2022.1022999

Zhang, P., Cao, B., & Li, F. (2022). The role of cognitive control in the SNARC effect: A review. *PsyCh Journal*, *11*(6), 792–803. https://doi.org/10.1002/pchj.586

Zorzi, M., Bonato, M., Treccani, B., Scalambrin, G., Marenzi, R., & Priftis, K. (2012). Neglect impairs explicit processing of the mental number line. *Frontiers in Human Neuroscience*, *6*, Article 125. https://doi.org/10.3389/fnhum.2012.00125

Downloaded from online.ucpress.edu/collabra/article-pdf/12/1/170661/952258/collabra_2026_12_1_170661.pdf by guest on 26 October 2026